



Titre: Caractérisation de l'offre et de la demande de transport dans un
Title: système de taxis électriques

Auteur: Arnould Geneste
Author:

Date: 2017

Type: Mémoire ou thèse / Dissertation or Thesis

Référence: Geneste, A. (2017). Caractérisation de l'offre et de la demande de transport dans
Citation: un système de taxis électriques [Mémoire de maîtrise, École Polytechnique de
Montréal]. PolyPublie. <https://publications.polymtl.ca/2600/>

 **Document en libre accès dans PolyPublie**
Open Access document in PolyPublie

URL de PolyPublie: <https://publications.polymtl.ca/2600/>
PolyPublie URL:

**Directeurs de
recherche:** Catherine Morency, & Martin Trépanier
Advisors:

Programme: Génie civil
Program:

UNIVERSITÉ DE MONTRÉAL

CARACTÉRISATION DE L'OFFRE ET DE LA DEMANDE DE TRANSPORT DANS UN
SYSTÈME DE TAXIS ÉLECTRIQUES

ARNAULD GENESTE

DÉPARTEMENT DES GÉNIES CIVIL, GÉOLOGIQUE ET DES MINES

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

MÉMOIRE PRÉSENTÉ EN VUE DE L'OBTENTION

DU DIPLÔME DE MAÎTRISE ÈS SCIENCES APPLIQUÉES

(GÉNIE CIVIL)

JUIN 2017

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Ce mémoire intitulé :

CARACTÉRISATION DE L'OFFRE ET DE LA DEMANDE DE TRANSPORT DANS UN
SYSTÈME DE TAXIS ÉLECTRIQUES

présenté par : GENESTE Arnould

en vue de l'obtention du diplôme de : Maîtrise ès sciences appliquées

a été dûment accepté par le jury d'examen constitué de :

M. LANGEVIN André, Ph. D., président

Mme MORENCY Catherine, Ph. D., membre et directeur de recherche

M. TRÉPANIÉ Martin, Ph. D., membre et codirecteur de recherche

M. BONNEL Patrick, Ph. D., membre

REMERCIEMENTS

Je souhaite remercier Catherine Morency qui m'a guidé tout au long de ce projet de recherche. Le dévouement quotidien dont elle fait preuve pour son métier et sa passion sont moteur de notre intérêt pour le transport. Je tiens à remercier Martin Trépanier dont la pertinence des conseils a orienté ma recherche, notamment sur l'aspect méthodologique.

Je remercie Jean-Simon Bourdeau et Hubert Verreault. Guidé par leur bienveillance, j'ai progressé tout au long de ce projet.

Je remercie Gabriel Lefebvre-Ropars et Jérôme Laviolette qui m'ont fourni de précieux conseils techniques durant ma recherche. Je remercie chaleureusement toute l'équipe et tous mes collègues de travail avec qui j'ai passé une belle année de travail.

Enfin je remercie l'École Nationale des Travaux Publics de l'État pour m'avoir donné l'opportunité d'étudier au Québec.

RÉSUMÉ

L'émergence de nouvelles technologies modifie petit à petit les conditions de notre mobilité. Parmi les nouvelles tendances en transport, nous pouvons citer le véhicule électrique qui, d'un certain point de vue, est un apport pour la mobilité durable. À Montréal, la société Téo Taxi opère déjà un réseau de taxis électriques sur une partie de l'île et cherche à s'implanter sur l'ensemble du territoire.

Avec l'apparition des technologies de l'information, les systèmes de transport par taxi sont aujourd'hui traçables par des données GPS. De précieuses données caractérisant les véhicules, les chauffeurs et les usagers du taxi deviennent accessibles. Une meilleure connaissance des systèmes de transport par taxi sert à optimiser leur utilisation à l'avantage des usagers qui sont aussi des citoyens soucieux de pouvoir se déplacer confortablement et sans entrave.

L'objectif principal de ce projet de recherche est le développement d'indicateurs de description et d'aide à l'évaluation de la performance d'un système de transport par taxi électrique. Afin d'atteindre cet objectif principal, les objectifs secondaires suivants sont poursuivis :

1. Définir les objets composant un système de transport par taxi électrique et les caractériser.
2. Automatiser le prétraitement des données GPS et de courses.
3. Développer une liste d'indicateurs qui serviront à donner un portrait global de l'utilisation du système de transport par les différents objets précédemment identifiés.
4. Construire ces indicateurs à partir des données prétraitées et en automatiser l'estimation.
5. Identifier des jours-types de demande intégrant les données temporelles et spatiales.
6. Proposer un algorithme de génération de l'offre en taxis électriques.

Une revue de littérature sur les taxis en général et les taxis électriques en particulier est présentée. Elle propose une définition d'un système de transport par taxi et fait l'état des lieux des politiques de législation encadrant son industrie. Elle propose un résumé de l'évolution historique du concept même de *taxi* et décrit les enjeux spécifiques aux taxis électriques. Les indicateurs de description des réseaux de taxis traditionnels et électriques qui sont dans la littérature sont présentés dans un tableau.

Des outils descriptifs ont été développés. Ce sont des indicateurs de description et d'aide à l'évaluation de la performance. Ils ont été construits pour la compagnie Téo taxi et contribueront à un second projet, plus vaste, qui a pour objectif d'établir un portrait global de l'offre et de la demande en taxi de toutes les compagnies de l'île de Montréal. Avant d'élaborer les indicateurs, une méthodologie reproductible de prétraitement de données a été développée. Un portrait global de l'utilisation de l'offre en taxis électriques de Téo Taxi sur l'île de Montréal a été effectué sur la période de juillet à décembre 2016. Cette analyse est la première de cette ampleur à avoir été faite pour le réseau de taxis électriques de Téo Taxi.

Pour résumer les faits saillants caractérisant l'offre et la demande, les chauffeurs rapportent près de 145\$ par jour. La majorité d'entre eux commencent à travailler le matin vers 6-7 h ou en début d'après-midi vers 13-15 h. À ce même moment, un grand nombre de chauffeurs finissent leurs quarts de travail, d'autres finissent plus tard vers 18-20 h. Quant au taux d'utilisation des véhicules disponibles, il tourne autour de 50% la journée et tend à diminuer vers 30% entre 2-5 h du matin. Le plus grand nombre de véhicules disponibles se trouve dans les mêmes secteurs de recensement que les sites de rechargement. Les usagers annulant leurs courses le font principalement parce que le temps d'attente est trop long ou qu'ils ont finalement changé d'avis. 50% des courses réalisées ont été demandées par 13% des usagers de Téo Taxi.

Afin d'harmoniser l'offre avec la demande, un travail de recherche a été effectué sur l'identification de jours-types de demande intégrant la donnée spatiale. Ce travail vise à terme à appuyer des compagnies comme Téo Taxi, qui seraient en mesure de décider du nombre de véhicules à envoyer tel jour à tel endroit. Enfin, un algorithme de génération de l'offre a été proposé. C'est un schéma explicatif et descriptif du fonctionnement d'un réseau de taxis électriques d'une part, et qui d'autre part permet de générer un nombre de véhicules et de bornes de recharge adaptées à une demande donnée.

ABSTRACT

The rise of new technologies is gradually changing the way we move. Among new tendencies in transportation, the use of electric vehicles can be seen as a tool to set up better sustainable cities. In Montreal, the Téo Taxi company is already operating a network of electric taxis and intends to cover the whole Montreal Island.

Thanks to information and communication technologies, taxi transportation systems are trackable via GPS. High valuable data describing vehicles, drivers and taxi users become accessible. A better understanding of the taxi transportation system helps optimize their use. Citizens, as end users, take advantage of this understanding to journey in a more fluid and comfortable manner.

The main purpose of this research project is to develop description indicators as well as indicators to assess the performance of a transportation system carried out by electric taxis. In order to achieve this main purpose, we strive at the same time to meet the following secondary targets:

1. Define and characterize all the objects which are part of an electric taxi transportation system.
2. Automate the pre-processing of GPS and fare data.
3. Develop a list of indicators based on the formerly characterized objects in order to give a global overview of the electric taxi transportation system.
4. Build these indicators based on pre-processed data and automate their estimate.
5. Identify patterns of demand. They should involve spatial and time-related data.
6. Build an algorithm generating the offer in electric taxis.

A literature review on the offer of taxis in general and the offer of electric taxis more specifically is presented. It suggests a definition of what is a taxi transportation system and presents a state of the art of the legal context of the taxi industry. A summary of the historical evolution of the concept of *taxi* is done and the specific challenges the electric taxis face are highlighted. Indicators describing the electric and conventional taxis's network have been extracted from the literature review and are listed in a table.

Descriptive tools have been developed. These are description indicators and performance assessment support indicators. Specifically designed for the Téo Taxi company, they should

support a larger project, which aims to set a global overview of offer and demand in taxi fares for all the taxi companies on the Montreal Island. Before developing the indicators, a replicable methodology of data pre-processing has been developed. Then a global overview of Téo Taxi's electric taxis use has been set up based on a period running from July to December 2016. This is the first time that a study has been conducted on such a large scale so as to describe Téo Taxi operational activity.

To sum up the prominent facts which depict offer and demand, taxi drivers earn nearly 145\$ a day. Most of them start working around 6:00-7:00 mornings or around 1:00-3:00 afternoons. During this afternoon period, many drivers end their work shift, while some others end later at about 6:00-10:00 p.m. 50% of the taxis are in use during the day time and tends to decrease to 30% between 2:00-5:00 a.m. Most of the listed available taxis were found close to recharging stations. Customers cancel their ride mostly because of an excessive waiting time or because they eventually changed their mind. 13% of Téo Taxi customers order 50% of its rides.

So as to balance offer and demand, a research has been conducted to identify day models demands which take the spatial dimension into account. This study aims to support taxi companies like Téo Taxi in defining how many vehicles to send on which locations depending on the day. Finally, an offer to generate algorithm has been developed. This is a descriptive and explicative diagram of the taxi operating system on the one hand, while, on the other hand, it also enables to match vehicles and recharging stations with a given demand.

TABLE DES MATIÈRES

REMERCIEMENTS	III
RÉSUMÉ.....	IV
ABSTRACT	VI
TABLE DES MATIÈRES	VIII
LISTE DES TABLEAUX.....	XII
LISTE DES FIGURES.....	XIII
LISTE DES SIGLES ET ABRÉVIATIONS	XVII
CHAPITRE 1 INTRODUCTION.....	1
1.1 Problématique.....	1
1.2 Objectifs	3
1.3 Structure du mémoire	3
CHAPITRE 2 REVUE DE LITTÉRATURE	5
2.1 Définition d'un système de transport par taxi.....	5
2.2 La régulation dans l'industrie du taxi	6
2.2.1 Réglementation au Québec.....	6
2.2.2 Régulation et libéralisation.....	7
2.3 Histoire de l'évolution du taxi	10
2.4 Évolution en Amérique du Nord et en Europe	24
2.5 Prévion de la demande.....	11
2.6 Spécificité des taxis électriques.....	12
2.5.1 Consommation d'électricité	12
2.5.2 Répartition des véhicules aux bornes de recharge	13
2.5.3 Localisation des bornes de recharge et répartition des courses entre les taxis.....	13

2.5.4	Véhicules autonomes.....	14
2.6	Indicateurs d'offre et de demande.....	15
CHAPITRE 3 MÉTHODOLOGIE GÉNÉRALE DE TRAITEMENT DE DONNÉES		21
3.1	Évolution de l'offre de service de transport par taxi électrique de Téo Taxi	21
3.1.1	Transformation de l'offre de mobilité dans l'industrie du taxi à Montréal	21
3.1.2	La compagnie de taxi électriques : Téo Taxi	22
3.1.3	Collecte des données	23
3.1.4	Évolution de la couverture spatiale	23
3.1.5	Évolution de l'offre et de la demande.....	24
3.2	Définition d'un système de transport par taxi électrique	28
3.2.1	Définition des objets.....	28
3.2.2	Description de la base de données.....	29
3.3	Méthodologie générale de prétraitement de données	33
3.4	Standardisation et nettoyages des données.....	35
3.4.1	Les règles de standardisation des données	35
3.4.2	Nettoyage des données : les règles d'exclusion	36
3.5	Table GPS : ajout de l'attribut identifiant le chauffeur du véhicule	38
3.5.1	Comment reconstituer l'attribut identifiant le chauffeur d'une table GPS ?.....	38
3.5.2	Calibration des zones de recharge	39
3.5.3	Méthodologie de reconstitution de l'attribut identifiant le chauffeur du véhicule	40
3.5.4	Limite de la méthode	42
3.6	Construction de la table GPS agrégée	43
3.6.1	Justification de la création de la table GPS agrégée aux 2 min.....	43
3.6.2	Présentation des attributs.....	43

3.7	Automatisation du traitement de données et optimisation	45
3.8	Indicateurs	47
3.8.1	Formes de visualisation utilisées	47
3.8.2	Présentation des indicateurs	48
3.8.3	Indicateurs d'offre.....	53
3.8.4	Indicateurs de demande	55
CHAPITRE 4	APPLICATION : UTILISATION DES INDICATEURS	60
4.1	Cadre d'étude	60
4.2	Application : Indicateurs d'Offre	62
4.3	Application : Indicateurs de Demande	70
4.4	Conclusion.....	91
CHAPITRE 5	IDENTIFICATION DE TYPOLOGIES DE JOURS DE DEMANDE	93
5.1	Construction d'une matrice de vecteurs.....	94
5.1.1	Types de vecteurs	94
5.1.2	Définition du territoire et de la temporalité de l'étude	95
5.1.3	Définition du type de déplacements	95
5.2	Segmentation	96
5.2.1	La méthode des k-moyennes	96
5.2.2	Dendrogramme et diagramme WCSS	96
5.2.3	Analyses	97
5.3	Analyses des résultats.....	98
CHAPITRE 6	ALGORITHME DE GÉNÉRATION DE L'OFFRE DE TAXIS ÉLECTRIQUES.....	107
6.1	Paramètres quantitatifs et qualitatifs	107
6.1.1	Modélisation du territoire d'étude	107

6.1.2	Les usagers	108
6.1.3	La flotte de taxi	108
6.1.4	Définition des seuils d'autonomie de la batterie.....	108
6.1.5	Les bornes de recharge	109
6.1.6	Les postes d'attente.....	110
6.2	Algorithme de génération de l'offre	110
6.2.1	Point de vue de l'agent usager	112
6.2.2	Point de vue de l'agent taxi électrique.....	112
6.2.3	Facteurs de décisions.....	113
6.2.4	Distribution des courses par l'opérateur	114
CHAPITRE 7 CONCLUSION		116
7.1	Contributions	116
7.2	Limitations	117
7.3	Perspectives	118
BIBLIOGRAPHIE		120

LISTE DES TABLEAUX

Tableau 2-1 : Revue des indicateurs.....	29
Tableau 3-1 : Informations données par la table des courses.....	42
Tableau 3-2 : Informations données par la table GPS.....	43
Tableau 3-3 : Tableau résumé des indicateurs fabriqués.....	62
Tableau 4-1 : Résultats du prétraitement des données - table de courses & table GPS.....	73
Tableau 4-2 : Tableau d'indicateurs d'offre.....	75
Tableau 4-3 : Tableau d'indicateurs de demande.....	84

LISTE DES FIGURES

Figure 2-1 : Niveau de charge d'une flotte de taxis électriques pour une journée	19
Figure 3-1 : Évolution de la couverture spatiale du réseau de Téo Taxi.....	24
Figure 3-2 : Evolution du nombre de chauffeurs	25
Figure 3-3 : Évolution du nombre de permis et de véhicules en fonction du temps	26
Figure 3-4 : Évolution du nombre de courses réalisées en fonction du temps.....	27
Figure 3-5 : Définition orientée-objet d'un système de transport par taxi électrique.....	28
Figure 3-6 : Diagramme de la base de données relationnelle de taxis électriques.....	32
Figure 3-7 : Méthodologie de prétraitement des données GPS et de courses de taxis.....	34
Figure 3-8 : Graphique résumant les règles de nettoyage de la table des courses	36
Figure 3-9 : Méthodologie de récupération de l'identifiant chauffeur	39
Figure 3-10 : Vue de la zone de recharge de l'aéroport. Vue cartographiée contre vue satellite. ..	40
Figure 3-11 : Méthodologie permettant de reconstituer l'attribut <i>driver_id</i> dans la table GPS	55
Figure 3-12 : Extrait d'une table GPS reconstituée	42
Figure 3-13 : Table GPS aux 2 minutes	45
Figure 3-14 : Illustration des formes de visualisation utilisées.....	47
Figure 4-1 : Débuts des quarts de travail pour une semaine moyenne.....	64
Figure 4-2 : Fins des quarts de travail pour une semaine moyenne	64
Figure 4-3 : Distribution des distances parcourues par jour par chauffeur et moyenne de la distance sans client - (Été + Automne).....	65
Figure 4-4 : Distribution des distances parcourue par jour par taxi et moyenne de la distance sans client - (Été + Automne).....	66
Figure 4-5 : Disponibilité et taux d'utilisation des taxis pour une semaine moyenne - (Été + Automne).....	67

Figure 4-6 : Densité horaire de disponibilité par secteur de recensement (h/km ²) pour une journée moyenne (tous types de services confondus) entre (a) 0-6h, (b) 6-9h, (c) 9-15h, (d) 15-18h, (e) 18-24h. - Été	68
Figure 4-7 : Densité horaire de disponibilité par secteur de recensement (h/km ²) pour une journée moyenne (tous types de services confondus) entre (a) 0-6h, (b) 6-9h, (c) 9-15h, (d) 15-18h, (e) 18-24h. - Automne	69
Figure 4-8 : Distribution des chauffeurs selon le nombre de courses effectuées par jour	71
Figure 4-9 : Distribution cumulée du nombre de courses en fonction du pourcentage cumulé des chauffeurs	72
Figure 4-10 : Distribution cumulée du revenu généré en fonction du pourcentage cumulé des chauffeurs	72
Figure 4-11 : Évolution du nombre de courses par heure-chauffeur travaillée - Été	73
Figure 4-12 : Évolution du nombre de courses par heure-chauffeur travaillée - Automne	73
Figure 4-13 : Comparaison des longueurs et vitesses moyennes pour une semaine moyenne (Été + Automne).....	74
Figure 4-14 : Distribution des longueurs de courses par période de la journée - Été	75
Figure 4-15: Distribution des longueurs de courses par période de la journée - Automne.....	75
Figure 4-16 : Distribution des durées des courses par période de la journée - Été	76
Figure 4-17 : Distribution des durées de courses par période de la journée - Automne	76
Figure 4-18 : Distribution cumulée des longueurs des courses en fonction du pourcentage cumulé des chauffeurs.....	76
Figure 4-19 : Distribution cumulée des durées de courses en fonction du pourcentage cumulé des chauffeurs	77
Figure 4-20 : Densité horaire d'origines et de destinations par secteur de recensement pour une journée moyenne (tous types de services confondus). - Été	79
Figure 4-21 : Densité horaire d'origines et de destinations par secteur de recensement pour une journée moyenne (tous types de services confondus). - Automne	81

Figure 4-22 : Taux de courses réalisées - Été.....	82
Figure 4-23 : Taux de courses réalisées - Automne	82
Figure 4-24 : Pourcentage de courses réalisées par types de services et de clients pour une semaine moyenne - (Été + Automne).....	83
Figure 4-25 : Pourcentage de courses annulées par types de services et de clients pour une semaine moyenne - (Été + Automne).....	83
Figure 4-26 : Densité horaire d'annulations par secteur de recensement pour une journée moyenne (tous types de services confondus) entre (a) 0-6h, (b) 6-9h, (c) 9-15h, (d) 15-18h, (e) 18-24h. - Été	85
Figure 4-27 : Densité horaire d'annulations par secteur de recensement pour une journée moyenne (tous types de services confondus) entre (a) 0-6h, (b) 6-9h, (c) 9-15h, (d) 15-18h, (e) 18-24h. - Automne.....	86
Figure 4-28 : Raisons des annulations pour une semaine moyenne (Été + Automne).....	Erreur ! Signet non défini.
Figure 4-29 : Nombre cumulé de courses annulées en fonction du temps d'annulation - (Été + Automne).....	87
Figure 4-30 : Distribution du nombre d'annulations par nombre d'usagers - (Été + Automne).....	88
Figure 4-31 : Longueur moyenne d'une course réalisée selon le temps d'attente de l'utilisateur - Été	89
Figure 4-32 : Longueur moyenne d'une course réalisée selon le temps d'attente de l'utilisateur - Automne	89
Figure 4-33 : Temps d'attente pour une semaine moyenne - Été	90
Figure 4-34 : Temps d'attente pour une semaine moyenne - Automne.....	90
Figure 4-35 : Distribution cumulée du nombre de courses en fonction du pourcentage cumulé d'usagers	91
Figure 5-1 : Méthode de construction de typologie de jours de demande	93
Figure 5-2 : Types de matrices réalisés	94
Figure 5-3 : Exemple de matrice de vecteurs <i>jour*sr</i> pour les déplacements originant	96

Figure 5-4 : Exemples de dendrogramme et de diagramme WCSS.....	97
Figure 5-5 : Diagramme indiquant quels résultats ont été ou n'ont pas été concluants.....	98
Figure 5-6 : Exemple de matrice de vecteurs <i>jour*sr²</i> par groupe d'heures.....	100
Figure 5-7 : Évaluation du nombre de clusters (matrice de vecteurs <i>jours</i>).....	100
Figure 5-8 : Pourcentage de déplacements quotidiens par heure de la journée, par type de jour	101
Figure 5-9 : Types de jours dominants par jour de l'année 2016 (vecteur <i>jours</i>).....	101
Figure 5-10 : Évaluation du nombre de clusters (matrice de vecteurs <i>jour*SR d'origine</i>)	103
Figure 5-11 : Pourcentage de déplacements quotidiens par heure de la journée, par type de jour	103
Figure 5-12 : Pourcentage de déplacements quotidiens par heure de la journée, pour les clusters 1 et 4.....	103
Figure 5-13 : Évaluation du nombre de clusters (<i>jour*paire OD</i> par groupe d'heures).....	104
Figure 5-14 : % de déplacements quotidiens par heure de la journée, par type de jour (<i>jours*sr²</i> par groupe d'heures)	104
Figure 5-15 : Types de jour dominants par jour de l'année 2016 (<i>jours*sr²</i> par groupe d'heures)	105
Figure 5-16 : Clusters dominants pour les secteurs d'origines et de destinations	105
Figure 6-1 : Exemples de spécificités caractérisant chaque zone du territoire	107
Figure 6-2 : Description des seuils de la batterie électrique	108
Figure 6-3 : Description orientée agent d'un réseau de taxis électriques	111

LISTE DES SIGLES ET ABRÉVIATIONS

DBSCAN	Density-Based Spatial Clustering of Applications with Noise
GES	Gaz à Effet de Serre
GPS	Global Positioning System (ou Système Mondial de Positionnement)
OD	Origines et Destinations
STM	Société de Transport de Montréal
SQL	Structured Query Language (ou Langage de Requête Structurée)
VAP	Véhicule Autonome Partagé
VLS	Véhicule en Libre-Service
VTC	Véhicule de Tourisme avec Chauffeur
WCSS	Within Cluster Sum Square (ou Variance Intra-Cluster)

CHAPITRE 1 INTRODUCTION

L'émergence de nouvelles technologies ainsi que des nécessités de transformations structurelles de nos modes de vie urbains contemporains, liées aux défis posés par le développement durable, modifient petit à petit les conditions de notre mobilité. Des tendances continuent à se confirmer comme l'autopartage ou le vélopartage, d'autres modes cherchent à se réaffirmer comme la marche ou le vélo individuels. Les nouvelles technologies comme la télécommunication soutiennent entre autres l'autopartage et le covoiturage en favorisant la mise en contact des usagers de ces modes. Ou encore, le traitement informatique des données numériques permet l'émergence du véhicule autonome.

Parmi les nouvelles tendances en transport, nous pouvons citer le véhicule électrique. La littérature technique accueille favorablement son arrivée au regard de son apport pour le développement durable (Lu et al., 2012; Meunier, 2012, 2015; Mierlo, 2013; Tang, 2016). En effet, roulant à l'électricité, un tel véhicule n'émet plus de gaz à effet de serre. Cependant, si nous nous penchons sur son cycle de vie complet, nous observons la persistance d'impacts environnementaux qui comportent leur lot de défi à relever (Hawkins et al., 2013; Mierlo, 2013). De ce fait, le véhicule électrique ne peut pas être l'ultime solution aux problèmes écologiques du transport, mais une solution à explorer pour tenter de construire durablement une société. À Montréal, cette solution prend forme avec l'arrivée d'un parc de taxis électriques. La société Téo Taxi a lancé un service de taxi opérant uniquement avec des véhicules électriques sur l'île de Montréal.

1.1 Problématique

L'étude du taxi en tant que moyen opérationnel de transport est un champ de recherche qui est resté longtemps peu abordé. La recherche portant sur la législation de l'industrie du taxi est plus ancienne que la recherche opérationnelle. Avec l'apparition des technologies de l'information, les systèmes de transport par taxi sont aujourd'hui traçables par des données GPS et il est possible d'obtenir de précieuses données caractérisant les véhicules, les chauffeurs et les usagers du taxi. Cependant, même si la recherche sur les taxis s'intensifie de plus en plus depuis ces dernières années, encore peu de contributions exploitent leurs données pour mieux comprendre les enjeux opérationnels de cette industrie.

La gestion opérationnelle de l'offre d'une compagnie de taxi porte sur une connaissance approfondie de sa flotte de véhicules, de son équipe de chauffeurs et de leur comportement au fil des jours. Sont-ils souvent disponibles ou souvent en course ? Sont-ils concentrés sur le territoire ou dispersés ? Mais la gestion porte tout aussi bien sur l'étude de la demande car il s'agit de la satisfaire au mieux. Combien de courses par jour sont commandées ? A quelle heure de la journée y en a-t-il le plus et où sont-elles localisées ? Le portrait d'utilisation d'une flotte de taxis peut faire l'objet d'une analyse systématique qui est aujourd'hui rendu possible par le traitement informatique des données émanant des nouvelles technologies de l'information et de la communication.

Les résultats offerts par ce portrait ont un pouvoir descriptif et explicatif. Ils peuvent notamment permettre au planificateur de connaître les faits saillants qui caractérisent les systèmes de transports par taxi et quelle utilisation est faite de ce mode dans la gamme d'offres de transport urbain. Une meilleure connaissance de ces systèmes peut aussi servir à optimiser leur utilisation, à l'avantage des usagers qui sont des citoyens soucieux de pouvoir se déplacer confortablement et sans entrave. Ils peuvent de plus permettre à l'opérateur de connaître les limites de la gestion opérationnelle de son offre et d'essayer de les repousser. En effet, un portrait global de l'utilisation de son réseau de taxi pourrait aider le gestionnaire à évaluer la performance de son système de transport.

Dans ce projet, il est entendu par « performance », la performance opérationnelle et non pas la performance économique. Quand la seconde traite de la maximisation du revenu pour le gestionnaire, la première traite de l'adéquation entre l'offre et la demande. Les courses commandées sont-elles toutes satisfaites ? Où se trouvent les courses annulées ? Combien de courses sont réalisées par heure de travail des chauffeurs ? Telles sont les questions auxquelles peut permettre de répondre une analyse descriptive d'un réseau de taxis, aidant alors le gestionnaire dans ses politiques de performance tant économique qu'opérationnelle.

Cependant, à l'essor de l'ère du numérique, le développement de méthodes d'exploitation des bases de données générées par l'industrie du taxi reste encore à faire. C'est l'objet de ce projet de recherche que d'exploiter les données d'un système de transport par taxi électrique.

Le traitement des données issues des nouvelles technologies peut de plus contribuer au développement de la modélisation numérique de l'offre et de la demande. Existe-t-il des jours

types de demande ? Est-il possible de spatialiser la demande? Mais aussi, quels impacts l'offre a-t-elle sur la demande et réciproquement ? Où placer stratégiquement les sites de recharge des taxis électriques ? Et combien faut-il de bornes par site ? La modélisation de l'offre et de la demande est un outil qui peut contribuer à répondre à ces questions.

1.2 Objectifs

L'objectif principal de ce projet de recherche est le développement d'indicateurs de description et d'aide à l'évaluation de la performance d'un système de transport par taxi électrique. De tels indicateurs permettent de réaliser un portrait global de l'utilisation de l'offre en taxis électriques à Montréal. Ils sont construits à partir de deux sources de données : une table GPS et une table de courses, la première caractérisant l'offre et la seconde la demande. Les indicateurs de ce projet ont été construits pour la compagnie Téo Taxi et contribueront à un second projet, plus vaste, qui a pour objectif d'établir un portrait global de l'offre et de la demande en taxi de toutes les compagnies de l'île de Montréal.

Afin d'atteindre l'objectif principal, les objectifs secondaires suivants sont poursuivis :

1. Définir les objets composant un système de transport par taxi électrique et les caractériser.
2. Automatiser le prétraitement des données GPS et de courses.
3. Développer une liste d'indicateurs qui servira à donner un portrait global de l'utilisation du système de transport à l'aide des différents objets précédemment identifiés.
4. Construire ces indicateurs à partir des données prétraitées et en automatiser l'estimation.

Deux objectifs secondaires ont été poursuivis en parallèle :

5. Identifier des jours-types de demande intégrant les données temporelle et spatiale.
6. Construire un algorithme de génération de l'offre en taxis électriques.

1.3 Structure du mémoire

Ce mémoire se compose de sept chapitres. Un chapitre introductif, un second portant sur la revue de littérature, le troisième porte sur la méthodologie générale développée pour ce projet, le

quatrième présente les résultats obtenus, l'identification des jours-types de demande est le cinquième chapitre, le sixième présente un algorithme de génération de l'offre. Enfin, le septième conclut cette recherche en donnant les perspectives possibles de ce projet.

À la suite de cette introduction, le second chapitre présente une revue de littérature qui fait l'état des lieux sur les politiques de législation de l'industrie du taxi, sur l'étude de la demande, sur l'évolution du concept de *taxi* et enfin sur la spécificité des taxis électriques.

Le troisième chapitre présente la méthodologie de ce projet. La définition des objets d'un système de transport par taxi électrique est faite. La méthodologie générale de prétraitement de données est exposée. Elle consiste en une étape de standardisation de la structure des tables des données GPS et de courses ainsi que d'un nettoyage de celles-ci. Une méthode de reconstitution de l'identifiant chauffeur de la table GPS est présentée. Suite à ce procédé, la construction d'une table GPS agrégée aux 2 minutes est exposée. La liste des indicateurs développés à partir des données prétraitées est décrite. Ces indicateurs portent sur l'offre et la demande d'un système de transport par taxi électrique.

Le chapitre quatre résulte du chapitre méthodologique précédent. Il présente l'analyse globale de l'utilisation de l'offre de transport de Téo Taxi entre juillet et décembre 2016 dans le territoire de desserte sur l'île de Montréal. Les différents indicateurs développés au chapitre précédent sont exploités afin d'illustrer une manière de les mettre en pratique.

Le cinquième chapitre porte sur l'identification de jours-types de demande. Une segmentation sur une matrice de vecteurs de demande permet de déterminer un ensemble fini de jours-types. Une première segmentation est faite en ne prenant en compte que la donnée temporelle. La donnée spatiale est ensuite ajoutée afin d'essayer de trouver des jours-types en fonction du jour de l'année et de sa localisation sur le territoire.

Le chapitre six présente un algorithme de génération de l'offre. Ce dernier sert d'outil explicatif pour comprendre le fonctionnement d'un système de transport par taxi électrique. Mais sa finalité est de donner les nombres de véhicules et de bornes nécessaires pour satisfaire une demande donnée.

CHAPITRE 2 REVUE DE LITTÉRATURE

Ce chapitre est dédié à la revue de littérature portant sur le taxi en général et le taxi électrique en particulier. Une définition de ce qu'est un système de transport par taxi est d'abord donnée. Ensuite, la régulation dans l'industrie du taxi est abordée et l'histoire de son évolution est tracée. Un état de l'art portant sur la prévision de la demande est proposé, ainsi qu'un autre portant sur les caractéristiques spécifiques aux taxis électriques (consommation d'électricité, problèmes de répartition aux bornes de recharge, localisation de ces dernières). Une revue d'indicateurs en relation avec l'offre et la demande de taxis a été réalisée, un tableau synthétique en résulte.

2.1 Définition d'un système de transport par taxi

Les systèmes traditionnels de transport par taxi sont composés d'une flotte de véhicules (les *taxis*), de conducteurs de véhicules (les *chauffeurs*), de *postes d'attente* et d'*intermédiaires* en service. Lorsque la flotte est composée de véhicules électriques, il faut ajouter les *bornes de recharge*. Un système de transport par taxi est une offre répondant à une demande particulière. La demande est composée des *usagers*. Quand l'offre satisfait la demande, c'est-à-dire quand un chauffeur de taxi transporte un usager de son origine jusqu'à sa destination, une *course* est réalisée. Il s'agit dans ce paragraphe de définir chacun de ces objets.

Un *taxi* est un véhicule ayant le rôle de conduire une ou plusieurs personnes d'un point d'origine à un point de destination. Le déplacement effectué est ainsi appelé une *course*. La course est tarifiée en fonction de la distance parcourue et du temps passé dans le véhicule. Les personnes dont les déplacements sont satisfaits sont appelées les *usagers* ou les *clients*. Un usager peut être ou ne pas être fidèle à une seule compagnie de taxi. Les taxis sont conduits par des *chauffeurs*. Ces derniers doivent posséder un *permis chauffeur* octroyé par le bureau du taxi (dans le cas de Montréal) s'ils veulent pouvoir conduire un taxi. Les chauffeurs de taxi peuvent travailler en tant que travailleurs autonomes ou salariés d'une compagnie de taxi. Quel que soit leur statut, ils ne peuvent conduire un véhicule qu'à condition que ceux-ci soient en possession d'un *permis de propriétaire de taxi* qu'ils ont obtenu ou qu'un propriétaire leur a confié. Quand ils sont travailleurs autonomes, ils doivent acheter ou louer leur véhicule et leur permis. Pour augmenter leur bénéfice, les chauffeurs se regroupent autour d'un centre de répartition des courses :

l'intermédiaire. Pour qu'une entreprise ou association deviennent un intermédiaire, elle doit nécessairement posséder un *permis d'intermédiaire*.

Quand un taxi est en attente d'un client, il a le choix entre parcourir les rues achalandées ou attendre à un poste d'attente. Un grand nombre de postes d'attente est disposé sur le territoire d'une ville et à des endroits stratégiques et achalandés ; ceux-ci peuvent être privés ou publics. Un poste d'attente est donc un endroit, souvent repéré à l'aide d'un panneau, où les chauffeurs garent leurs véhicules et attendent qu'un client leur fasse signe. Ce signe peut se manifester de plusieurs manières : les trois principales manières sont le hélage, la réservation par appel téléphonique et la réservation par application mobile. Le rapport que l'industrie du taxi entretient aujourd'hui avec l'émergence des applications mobiles a été approfondi par Harding (2016).

Le hélage est la manière la plus directe : un usager a besoin d'un taxi et hèle parfois à grand cri ou avec de grands gestes un taxi passant ou garé à un poste d'attente non loin. La réservation d'une course par téléphone ou application mobile se fait à l'aide d'un intermédiaire. Quand l'utilisateur fait sa réservation, l'intermédiaire assigne la course commandée à un chauffeur disponible. En ce qui concerne Téo Taxi, la compagnie est propriétaire des véhicules et locataire des permis de taxi.

Si les véhicules utilisés auparavant étaient uniquement des véhicules à essence, ils peuvent dorénavant être des véhicules roulant à l'électricité. Étant équipés d'une batterie, les taxis électriques doivent régulièrement retourner à une borne de recharge où ils seront rechargés. À Montréal, de nombreuses bornes de recharge sont possédées par Téo Taxi. D'autres sont publiques, d'autres encore sont privées - comme celles appartenant à la compagnie Tesla. Il existe plusieurs types de bornes correspondant à différentes vitesses de charge. Le concept de *site de bornes* a émergé durant ce projet de recherche. Un site de bornes est une zone de recharge où se trouvent plusieurs bornes.

2.2 La régulation dans l'industrie du taxi

2.2.1 Réglementation au Québec

Le Québec réglemente l'industrie du taxi par la *loi concernant les services de transport par taxi*. Elle assure la qualité des services offerts par les compagnies de taxi, une gestion de l'offre qui

tient en compte les besoins de la population et sa sécurité, mais aussi afin de réglementer l'activité des intermédiaires. Un intermédiaire est :

toute personne qui fournit, par tout moyen, à des titulaires d'un permis de propriétaire de taxi ou à des titulaires d'un permis de chauffeur de taxi des services de publicité, de répartition de demandes de services de transport par taxi ou d'autres services de même nature. (Gouvernement du Québec, 2016)

Pour le gouvernement du Québec, un taxi est : "un véhicule automobile exploité en vertu d'un permis délivré en application de la Loi concernant les services de transport par taxi" (Gouvernement du Québec, 2016). Il est clarifié qu'un "permis de propriétaire de taxi autorise son titulaire à posséder un seul taxi". Cependant l'arrivée de Téo Taxi à Montréal a modifié la réglementation, car les législateurs ajoutent :

Malgré le premier alinéa, lorsque l'automobile servant au transport par taxi est mue entièrement au moyen de l'énergie électrique, la Commission des transports du Québec peut autoriser le titulaire du permis de propriétaire de taxi à posséder le nombre d'automobiles supplémentaires mues entièrement au moyen de l'énergie électrique qu'elle détermine pour s'assurer que le titulaire du permis puisse continuer d'offrir des services pendant le temps de la recharge. (Gouvernement du Québec, 2016)

La réglementation entre dans une politique de régulation de l'activité de l'industrie du taxi. Cependant d'autres courants politiques promeuvent la dérégulation - ou la libéralisation - du marché.

2.2.2 Régulation et libéralisation

Le service de transport traditionnel par taxi est depuis longtemps un service réglementé par la loi, ceci afin d'assurer une qualité et un niveau de service suffisant, et de neutraliser les comportements anticoncurrentiels. Il faut distinguer 3 types de régulations : la régulation de la qualité et du niveau de service, ainsi que la régulation de la conduite du marché. Cette dernière régulation régleme les comportements concurrentiels des opérateurs (Salanova, 2011 ; Harding, 2016). Contrôler le nombre de taxis sur la route permet entre autres d'assurer un digne revenu aux chauffeurs et de contrôler le niveau de pollution et de congestion en ville. Le nombre de taxis, le nombre de permis par zone, le tarif d'une course et le nombre de permis autorisés sont les 3 objets sur lesquels agissent les législateurs pour la régulation de l'industrie.

Dans les années 60-70, un débat entre les chercheurs émerge pour savoir s'il est mieux de faire une politique libérale ou une politique de régulation (Darbéra, 2007). Carins et Liston-Heyes, (1996) démontrent que, parce que l'industrie du taxi ne se prête pas à la compétition, la dérégulation des tarifs des courses et du nombre de concurrents n'est pas un choix optimal. De nombreuses villes états-uniennes dans les années 80 ont essayé l'expérience libérale, mais ont finalement fait marche arrière suite à de nombreuses crises de l'industrie. Ils assurent aussi que la mise en place de permis de taxis régule la qualité des relations entre chauffeurs et donc du service de transport par taxis (Cairns et al., 1996).

Certains auteurs s'accordent sur le fait qu'il existe deux marchés dans l'industrie du taxi : le marché correspondant aux taxis hélés et le second correspondant aux taxis réservés (Bekken et al., 2007; Darbéra, 2010). En Amérique du Nord, si l'absence de régulation du nombre de taxis sur le marché du taxi hélé conduit à un nombre trop important de véhicules en service, néanmoins une régulation trop serrée a des chances de générer un manque de taxis et empêche de ce fait de satisfaire la demande sur les deux marchés (Schaller, 2007).

De nos jours encore, la régulation de l'industrie du taxi est en débat. Une forte critique argumentée a été publiée par Salanova et al., 2011. Selon ces auteurs, depuis la crise économique de 2008, la régulation traditionnelle provoque une distorsion du marché (augmentation du nombre de chauffeurs, augmentation du prix des permis). La dérégulation, toujours selon ces auteurs, apporte de nombreux avantages : baisse du temps d'attente, baisse du niveau des prix, augmentation de la satisfaction de l'utilisateur. La dérégulation du niveau de service entraîne le plus souvent une augmentation de sa qualité et par conséquent assure un confort et un service minimum de base au chauffeur et à l'utilisateur (Salanova et al., 2011).

Il existe des recherches au sujet du potentiel d'utilisation des applications mobiles pour l'opérationnalisation d'une flotte de taxis (Ruyuan Li, 2012 ; Chen, 2014). Il est à noter que l'introduction de l'application mobile dans cette industrie l'a transformée et que pourtant, comme le signale Harding (2016) encore peu d'attention est accordée à l'impact des nouvelles technologies sur la régulation du taxi. Ceci probablement parce que l'usage d'applications mobiles est un phénomène récent. Aux États-Unis, 61% des utilisateurs du téléphone possédaient un téléphone intelligent début 2013. Selon l'auteur, la régulation ne devrait pas s'attacher à réglementer l'utilisation des applications dans l'industrie du taxi, mais plutôt les conséquences

que cela risque d'engendrer à savoir les collusions et les monopoles acquis par la domination numérique maintenant rendus possibles grâce aux applications mobiles (Harding et al., 2016).

2.2.3 Concurrence entre taxis et VTC

L'industrie du taxi connaît à ce jour une remise en question liée à l'émergence de la concurrence des voitures de tourisme avec chauffeur (ou VTC), autrefois appelées *voitures de petite remise*. Il existe deux marchés pour les taxis : le marché des courses héliées et le marché des courses réservées. C'est sur ce dernier marché que les VTC accroissent notamment la concurrence, utilisant les applications mobiles pour faire entrer en relation directe le chauffeur et l'utilisateur. Les applications permettent le maraudage électronique et concurrencent les centrales d'appels téléphoniques des intermédiaires de taxi (Darbéra, 2014b). L'arrivée de cette nouveauté technologique a permis l'essor de compagnies comme UberX qui, faisant face à un vide juridique dans la législation de l'industrie du taxi, a pu se développer au détriment des services de taxis. Une étude réalisée à San Francisco à l'aide d'une technique de sondage montre que le VTC a un point fort non négligeable par rapport au taxi : l'utilisateur qui commande une course attend moins longtemps et le temps d'attente est plus fiable (Lisa Rayle et al., 2016). Une étude réalisée sur une base de données GPS comprenant 18% des taxis enregistrés de la ville de Shenzhen en Chine entre janvier 2013 et novembre 2015 montre comment le phénomène des VTC s'est très rapidement propagé et s'est stabilisé en 2015 dans cette ville (Yu (Marco) Nie, 2017). L'étude montre aussi les points suivants : les taxis arrivent à concurrencer les VTC durant les heures de pointe dans les zones de fortes densités de population et les VTC aggravent la congestion urbaine même si cette aggravation est légère à Shenzhen. L'auteur recommande enfin la régulation du marché des courses héliées. Certains chercheurs s'accordent sur la nécessité de l'adaptation de l'industrie du taxi qui doit prendre en main la technologie numérique (Darbéra, 2014a; Geradin, 2015). Darbéra (2014b) montre que même si en France la polémique entre taxis et VTC se poursuit, en Angleterre au contraire les centrales d'appels téléphoniques sont devenues surannées par rapport aux réservations par application mobile.

2.3 Histoire de l'évolution du taxi

Dans ce paragraphe, l'histoire de l'évolution du taxi sera abordée en essayant de tenir compte de son évolution en Europe et en Amérique du Nord. Les premières traces du taxi sont repérables à Paris et à Londres. Puis quand le taxi commence à se motoriser, il est possible de suivre son évolution en Amérique du Nord.

Dans son origine, l'histoire du taxi s'articule avec l'histoire du transport public. En effet, qu'est-ce que le taxi si ce n'est un véhicule partagé conduisant une ou plusieurs personnes d'un point d'origine à un point de destination ? Ainsi les chaises porteuses, les voitures hippomobiles, puis les fiacres au 17^e siècle répondent à cette définition. C'est avec la naissance de l'automobile que naît le taxi comme nous le connaissons aujourd'hui (Taxi Lille, 2017).

C'est à Paris puis à Londres au début du 17^e siècle que nous trouvons les premières émergences du concept du taxi. Blaise Pascal invente en 1662 le service hippomobile sur route fixe à Paris. Les services de transports publics sont, au début, seulement réservés aux bourgeois (Vuchic, 2007). La trace des premiers postes d'attente remonte à 1654 quand ils sont mentionnés dans un document juridique de Londres. Du 17^e au 19^e, les véhicules étaient donc tirés par des chevaux, évoluaient sur des trajets fixes ou faisaient du porte-à-porte et étaient soit réservés à l'avance par les usagers, soit hélés. Le dernier véhicule hippomobile en date avant l'apparition des véhicules motorisés était une voiture à deux roues où le passager arrière surplombait le conducteur. Ce modèle était répandu en Angleterre et en Amérique (Ray, 2010).

C'est à partir de 1890 que les premiers moteurs électriques apparaissent. Mais ils sont très vite abandonnés au profit des moteurs à pétrole et au diesel qui sont des énergies plus rentables (Taxi Lille, 2017) faisant disparaître du même coup les organisations qui les entretenaient comme la London Electric Cab Company à Londres ou l'Electric Carriage and Wagon Company aux États-Unis. Cette dernière compagnie dura au moins 10 ans, de 1897 à 1907 quand les premiers véhicules à pétrole furent importés de France, où le premier véhicule vit le jour par le manufacturier Prunel. Ce dernier introduisit ses premiers modèles à Paris et à Londres en 1903 (De Cooper, 2010).

En 1909 le mot *taxi* apparaît suite à la mise en place du taximètre sur les véhicules, grâce à l'astucieux Kratz-Boussac. Au début, les voitures étaient nommées les "taximètres automobiles",

nom qui s'est ensuite transformé en "taxis" (Taxi De France, 2017). Ce n'est que récemment avec la popularisation du téléphone que les courses réservées ont augmenté, c'est-à-dire dans les années 60 pour l'Amérique du Nord et 70 pour l'Europe.

Les permis de taxis apparaissent en Angleterre au 17^e siècle ainsi que la tendance à la régulation. C'est durant la Grande Dépression des années 1930 qu'aux États-Unis vint le besoin de réguler le marché du taxi. L'objectif premier était de protéger l'intérêt de l'industrie plutôt que de faire prévaloir le service public rendu par les taxis (Zegras, 2011).

L'industrie du taxi connaît une révolution intérieure liée à l'apparition des nouvelles technologies du numérique. Ainsi, fondamentalement, le concept de "taxi" n'a pas changé au cours des siècles, mais sa forme matérielle, elle, s'est transformée au fil de l'histoire. Autrefois des chaises porteuses pour les bourgeois, les taxis sont aujourd'hui devenus des automobiles motorisées démocratisées. Les technologies du numérique continuent à transformer la relation entre l'utilisateur et le chauffeur. Au 21^e siècle, l'ère du véhicule électrique est relancée. Une première flotte de taxis électriques a vu le jour en 2015 à Montréal, Québec.

2.4 Prédiction de la demande

L'exploitation des données issues des technologies de l'information et l'informatisation des modèles permet aux chercheurs de se pencher sur le problème de la prédiction de la demande en taxi. La prédiction de la demande se fait dans le temps et dans l'espace. En effet, dans le cas d'un réseau de taxis, il s'agit de savoir où sur le territoire et quand dans le temps vont avoir lieu les prochaines courses. Prédire la demande permet d'organiser l'offre de manière optimale pour satisfaire le maximum d'utilisateurs.

À partir de l'historique des données, des modèles de prédiction à court-terme ont été proposés. Le modèle de Damas (2012) agrège à la fois le modèle ARIMA et celui de Poisson. Il peut prédire à un horizon de 30 min avec un taux de précision de 78%, c'est-à-dire que 78% des courses prévues ont réellement eu lieu. Moreira-Matias et al. (2013) propose un nouveau modèle en s'appuyant cette fois-ci sur un modèle d'apprentissage basé sur des données GPS. Celui de Lee et al. (2012) utilise la méthode des réseaux de neurones artificiels pour prédire la demande d'un système d'autopartage électrique. Ils arrivent à obtenir moins de 4% d'erreur pour 92,9% des prévisions ayant un horizon d'1h, et 77,7% pour celles ayant un horizon de 2h. Yunxuan et al. (2017)

proposent un nouveau modèle de prévision à court-terme, Wave SVM, qui est plus performant que ceux qu'ils ont trouvés en faisant un état de l'art. Denis et al. (2017) comparent trois algorithmes : les algorithmes de Markov, de Lempel-Ziv-Welch et des réseaux de neurones, notant que dans leur étude l'algorithme de Markov est meilleur que celui des réseaux de neurones.

D'autres méthodes de prévision de la demande sont réalisées à l'aide des méthodes de segmentation. Chang et al. (2010) utilisent trois méthodes de segmentation différentes pour prévoir la concentration de la demande dans l'espace, dans le temps et en fonction de la météo et les compare. Lu Zhang et al. (2016) utilisent l'algorithme DBSCAN pour segmenter les données et trouver des points de concentration de la demande. DBSCAN ne gérant pas très bien les données massives, il utilise les algorithmes Grid et Kd-tree qui sont complémentaires. Parmi les méthodes de segmentation, la méthode des k-moyennes est la plus populaire, car elle est simple à mettre en œuvre, elle est efficace et assure de bons résultats quand elle se fonde sur des données empiriques. Sa popularité montre aussi la difficulté qu'il y a à inventer un algorithme de segmentation de données (Jain, 2009).

2.5 Spécificité des taxis électriques

L'attention portée sur les véhicules électriques dans la littérature est très récente cependant le sujet reste encore peu documenté. Depuis 2015, le nombre d'articles augmente rapidement. Les articles produits viennent notamment d'Amérique du Nord ou d'Europe, mais aussi de la Chine. Les Chinois ont chez eux des flottes de véhicules hybrides et électriques. En 2016, un grand nombre d'articles documentant leurs projets de développement est sorti. Une méthode pour augmenter le taux d'utilisation du moteur électrique des véhicules hybrides à Pékin a été proposée (Meng Li, 2016). Une analyse du cycle de vie des véhicules hybrides, à Pékin encore, montre qu'entre 2016 et 2020 il y aurait une réduction de GES (Tang, 2016). Une analyse de faisabilité de l'implantation d'une flotte de taxis électriques, en insistant sur les points de vue financier et environnemental, a été réalisée pour Séoul (Meng Li, 2016).

2.5.1 Consommation d'électricité

Maia et al. (2011) ont simulé la consommation d'électricité de véhicules électriques. Ils ont proposé un modèle physique permettant de calculer la consommation, validé à l'aide de circuits tests. Ils montrent que la pente d'un terrain a un impact sur la performance du véhicule, puisque

par exemple les véhicules peuvent recharger leur batterie quand ils freinent en descendant. L'utilisation de la grille électrique résidentielle par les véhicules électriques en Europe pose un problème auquel des chercheurs ont tenté de répondre. Afin d'éviter la surinstallation de bornes de chargement rapide et pour permettre une rentabilité économique du réseau électrique, la gestion de l'utilisation de ce dernier par un quartier résidentiel devra être harmonisée dépendamment de l'utilisation des véhicules électriques. Ceci permettrait alors une bonne gestion de la demande en électricité. Laissant alors la place aux pics d'utilisation résidentielle, tout en laissant la possibilité aux véhicules électriques de se recharger rapidement (Dyke et al., 2010).

2.5.2 Répartition des véhicules aux bornes de recharge

Le véhicule électrique a la particularité d'être dépendant de l'autonomie de la batterie. Cette donnée doit être prise en compte dans les problèmes d'opérationnalisation du taxi en centre urbain. Un très grand nombre de facteurs impacte la décharge de la batterie, nous pouvons citer la vitesse, l'accélération, le type de véhicule et la température. La gestion de l'autonomie du véhicule cherche à savoir où et quand un véhicule doit aller se recharger. Un modèle développé par Rocca (2016) permet de tendre vers une uniformisation de l'utilisation des stations dans le temps. Cependant, il oblige les taxis à être moins disponibles pour la demande et peut générer des files d'attente. En essayant de prendre en compte la demande future et le niveau de charge des véhicules, Lu et al. (2012) proposent un modèle qui tente de réduire le temps d'attente aux bornes de recharge. Ils prennent en compte trois cas : le cas où les stations sont des lieux où la batterie est rechargée, le cas où la batterie est seulement échangée et le cas où il existe des stations d'échange et de rechargement.

2.5.3 Localisation des bornes de recharge et répartition des courses entre les taxis

Dans la littérature, encore peu de modèles d'optimisation de la localisation des lieux d'implantation des bornes de recharges ont été proposés. Souvent les chercheurs essaient d'optimiser à la fois la localisation des lieux de recharge et la répartition des réservations des courses des usagers.

Une étude germano-polonaise (Maciejewski & Bischoff, 2014) a modélisé l'impact de l'utilisation d'une flotte de véhicules électriques pour un service de taxi. Aucun impact négatif n'a été

démontré, mais face à une demande accrue une flotte de taxis traditionnels peut fournir de meilleurs services. L'étude montre de plus qu'un nombre restreint de bornes de recharge à chaque station de recharge est suffisant. La localisation des bornes de recharge pour une flotte de taxis électriques doit être réalisée stratégiquement afin de pouvoir maximiser le revenu horaire des chauffeurs. Pour une semaine de données, la demande en rechargement varie de jour en jour, et d'heure en heure. La distance de service et la couverture du territoire desservi a un impact sur la localisation des bornes (Joonho et al., 2017). Un modèle d'optimisation combinant un modèle d'attribution de postes de recharge électrique avec délais de file d'attente et répartition des appels de demande en taxi a été proposé. Le but du modèle est de minimiser les points de relève et le temps de voyage des passagers, ainsi que de fournir l'emplacement optimal des lieux de recharges et d'allocation de chargeurs. Celui-ci est testé sur une flotte de 600 taxis de taxi-partage à Séoul (Park, 2014). Afin de déployer une flotte de taxis électriques dans Paris, les problèmes opérationnels de l'implantation de bornes de recharge ont été étudiés. Les problèmes soulevés sont les suivants : trouver la juste place des bornes de recharge électrique, trouver le juste nombre de ces bornes et trouver un algorithme de répartition de la demande entre les taxis en prenant en compte le besoin de rechargement des batteries électriques. Deux manières de trouver l'emplacement des bornes électriques ont été proposées, chacune appuyée par un modèle. Le premier modèle est un modèle de minimisation des distances entre la borne de recharge et la position des clients dans une zone géographique donnée. Le second modèle, plus concluant, propose de localiser les lieux d'implantation des bornes de recharge en fonction des zones géographiques qui couvrent un maximum de demande (Meunier, 2012, 2015).

2.5.4 Véhicules autonomes

Le taxi électrique est actuellement conduit par un chauffeur. Avec l'émergence de l'intelligence artificielle, des véhicules électriques sans chauffeur ont été imaginés : des taxis autonomes. Si de nombreuses études techniques ont été effectuées sur les véhicules autonomes, seulement quelques études sur les services potentiels rendus à la population urbaine ont été réalisées à ce jour. L'utilisation de taxis autonomes en libre-service pourrait amener à terme à une diminution du nombre de voitures privées particulières en centre urbain, réduisant l'empreinte écologique et améliorant la qualité de vie des résidents. Maciejewski (2016) a modélisé le remplacement des véhicules privés par des flottes de taxis autonomes de tailles différentes à Berlin. Le transport en

commun est pris en compte en imposant une contrainte de vitesse dépendante du temps sur chaque lien du réseau routier modélisé. Un algorithme d'assignation des taxis à la demande des usagers a été proposé et son utilisation montre qu'une flotte de 100 000 véhicules serait suffisante pour remplacer l'actuel parc automobile de Berlin avec un haut niveau de service. Le taxi autonome peut avoir pour autre nom : le véhicule autonome partagé (VAP). Chen et al. (2016) étudient la gestion de véhicules électriques et autonomes partagés pour une ville ressemblant à Austin. Ils ont effectué plusieurs simulations en se basant sur différents scénarios de localisation des bornes de recharge et différents types de flottes (à haute et basse autonomie, avec véhicules électriques seulement ou en partie). Certains résultats sont notables. La taille de la flotte est fortement dépendante du nombre de véhicules et du type de bornes de recharge utilisé. Tous les 130 véh-km, une flotte de VAP électriques est en mesure de remplacer 3,7 véhicules privés par VAP électrique en service pour une recharge à 240 Volt AC, et 5,5 véhicules privés tous les 320 veh-km. De plus, 95-98% de 10% de déplacements métropolitains sont satisfaits par les VAP électriques, pour un temps moyen d'attente entre 7-10 min par voyage pour les usagers. Une autre étude (Rigole, 2014) analyse l'impact d'un service de taxis autonomes à Stockholm. Un algorithme d'assignation des courses aux véhicules est développé, une analyse de cycle de vie est effectuée. En conclusion un service de voitures autonomes est capable de satisfaire autant de déplacements, mais en utilisant seulement moins de 10% du nombre de véhicules actuels, et une diminution de près de 90% du kilométrage parcouru. Cependant si les utilisateurs ne sont pas prêts à diminuer leur niveau de service (augmentation du temps d'attente, présence de passagers supplémentaires, etc.), les voitures autonomes risquent d'augmenter l'intensité de la circulation et l'empreinte environnementale du secteur du transport.

2.6 Indicateurs d'offre et de demande

Cette section concerne les indicateurs permettant de faire l'état descriptif de l'offre et de la demande d'un système de transport par taxi. Le tableau 2-1 ci-dessous présente les indicateurs trouvés dans littérature. Il a été originellement établi par Lacombe (2016) et est mis à jour ici dans le cadre de ce projet de recherche. Les auteurs mis en italique sont les auteurs ajoutés au tableau originel. L'ajout majeur à ce dernier concerne tous les indicateurs ayant trait aux véhicules électriques.

Tableau 2-1 : Revue des indicateurs

OBJETS	INDICATEURS	RÉFÉRENCES	
Course	Carte de chaleur des origines et destinations	<ul style="list-style-type: none"> • (Ferreira et al., 2013) • (Pelé et al., 2014a) 	<ul style="list-style-type: none"> • (Lei Zhang et al., 2012) • (Yifan Zhang, 2014)
	Nombre d'origines et de destinations	<ul style="list-style-type: none"> • (Ferreira et al., 2013) • (King et al., 2012) • (New York City Taxi & Limousine Commission, 2014) 	<ul style="list-style-type: none"> • (Savage et al., 2013) • (Yifan Zhang, 2014)
	Cartographie des origines	• (Maciejewski, 2016)	
	Nombre de courses	<ul style="list-style-type: none"> • (Bischoff et al., 2015) • (Ferreira et al., 2013) • (Maciejewski, 2014) • (Maciejewski, 2016) 	<ul style="list-style-type: none"> • (New York City Taxi & Limousine Commission, 2014) • (Pelé et al., 2014a) • (Savage et al., 2013)
	Nombre de courses en fonction de la vitesse de course	• (Rigole, 2014)	
	Nombre de courses en fonction de leur durée	• (Roy et al., 2011)	
	Distance et durée de courses	<ul style="list-style-type: none"> • (Bischoff et al., 2015) • (New York City Taxi & Limousine Commission, 2014) 	<ul style="list-style-type: none"> • (Pelé et al., 2014a) • (Savage et al., 2013) • (Dao-Zheng Zhang et al., 2014) • (Meunier, 2012)
	Vitesse de course	<ul style="list-style-type: none"> • (Pelé et al., 2014a) • (Savage et al., 2013) 	• (Dao-Zheng Zhang et al., 2014)
	Nombre de passagers par course	• (Savage et al., 2013)	
	Niveau de consommation (électrique) d'une course	• (Lu et al., 2012)	
Taxi	Distance et durée à vide	<ul style="list-style-type: none"> • (Bischoff et al., 2015) • (Jun, 2013) 	<ul style="list-style-type: none"> • (Pelé et al., 2014a) • (Dao-Zheng Zhang et al., 2014)
	Distance et durée totales parcourues par véhicule et pour la flotte entière	• (Rigole, 2014)	

Tableau 2-1 : Revue des indicateurs (suite)

OBJETS	INDICATEURS	RÉFÉRENCES
	Vitesse par heure de la journée	• (Roy et al., 2011)
	Temps en service	• (Pelé et al., 2014a)
	Nombre de courses	• (Pelé et al., 2014a)
	Taux d'occupation	• (Yifan Zhang, 2014) • (Rigole, 2014)
	Nombre de taxis disponibles	• (Bischoff et al., 2015) • (Ganti et al., 2011) • (Dao-Zheng Zhang et al., 2014) • (New York City Taxi & Limousine Commission, 2014) • (Pelé et al., 2014a)
	Niveau de charge	• (Maciejewski, 2014)
	Nombre de véhicules en charge par heure de la journée	• (Maciejewski, 2014)
	Durée passée selon les modes d'opération, par heure de la journée	• (Maciejewski, 2016)
	Classement des véhicules par leur part de temps passé à vide, par heure de la journée	• (Maciejewski, 2016)
	Nombre de véhicules par heure de la journée, par motif de déplacement	• (Roy et al., 2011)
Borne	Temps d'attente du chauffeur en fonction des types de bornes.	• (Lu et al., 2012)
	Temps d'attente en fonction du nombre de bornes	• (Meunier, 2012)
	Longueur de la file d'attente selon l'heure de la journée	• (Meunier, 2012)
	Taux d'occupation des sites de bornes selon l'heure de la journée	• (Meunier, 2012)
	Temps d'attente par site de bornes en fonction du niveau de demande.	• (Lu et al., 2012)

Tableau 2-1 : Revue des indicateurs (suite)

OBJETS	INDICATEURS	RÉFÉRENCES
Chauffeur	Temps travaillé	• (Dao-Zheng Zhang et al., 2014)
	Début et fin des quarts de travail	• (New York City Taxi & Limousine Commission, 2014)
	Revenu ou profit	• (Ferreira et al., 2013) • (Jun, 2013) • (New York City Taxi & Limousine Commission, 2014) • (Dao-Zheng Zhang et al., 2014) • (Yifan Zhang, 2014)
	Temps d'arrivée du chauffeur au client	• (Maciejewski, 2014)
	Ratio du temps des chauffeurs passé à conduire à vide par heure de la journée	• (Maciejewski, 2016)
Zone	Temps de transition	• (Lei Zhang et al., 2012)
	Nombre de transitions entre 2 régions	• (Lei Zhang et al., 2012)
	Étude spécifique d'une zone (ex. : aéroport)	• (Bischoff et al., 2015) • (New York City Taxi & Limousine Commission, 2014)
Client	Distance de marche entre l'origine et le centre du lieu (ou poste) d'attente des taxis	• (Dao-Zheng Zhang et al., 2014)
	Caractéristiques des clients (âge, sexe, etc.)	• (New York City Taxi & Limousine Commission, 2014)
	Temps et longueur de course par client	• (Rigole, 2014)
	Temps d'attente en fonction du niveau de demande	• (Maciejewski, 2014)
	Temps d'attente en fonction du nombre de bornes de recharge utilisé	• (Maciejewski, 2014)

L'objectif principal de ce projet est le développement d'indicateurs qui permettent de dresser le portrait de l'utilisation du taxi d'une compagnie de taxis électriques. Le tableau ci-dessus nous permet donc d'avoir un état des lieux d'indicateurs ayant permis la description et l'analyse d'un tel

type de service de transport. Le tableau est orienté objet : il distingue les différents objets d'un système de transport par taxi. Ceux-ci sont analysés à l'aide de différents indicateurs qui décrivent leur évolution dans l'espace et dans le temps.

L'objet *borne* a été ajouté afin d'énoncer les indicateurs de la littérature portant sur les bornes de recharge. Le nombre d'indicateurs de l'objet *client* a été enrichi, notamment en ce qui concerne les diverses manières de mesurer le temps d'attente. Pour l'objet *taxi*, l'indicateur *durée passée selon les modes d'opération, par heure de la journée* a été ajoutée. Il est entendu par "mode d'opération", les modes suivants : véhicule vide, véhicule en train d'arriver au client, véhicule occupé, véhicule en train de déposer le client.

La course est étudiée quantitativement. Sa vitesse, sa distance, sa durée, sa consommation énergétique sont aussi étudiées. Selon Pelé et Morency (2014), pour la flotte de Taxi Diamond à Montréal composée de 968 taxis en octobre 2011, la distance moyenne d'une course est de : 5,2 km (coef. var. : 117%), sa durée moyenne : 17,4 min (coef. var. : 98%), sa vitesse moyenne : 16,9 km/h (coef. var. : 60%). La demande peut aussi être décrite à l'aide de l'utilisateur du taxi (ou client), à l'aide de ses caractéristiques sociodémographiques de l'utilisateur, mais aussi en calculant son temps d'attente.

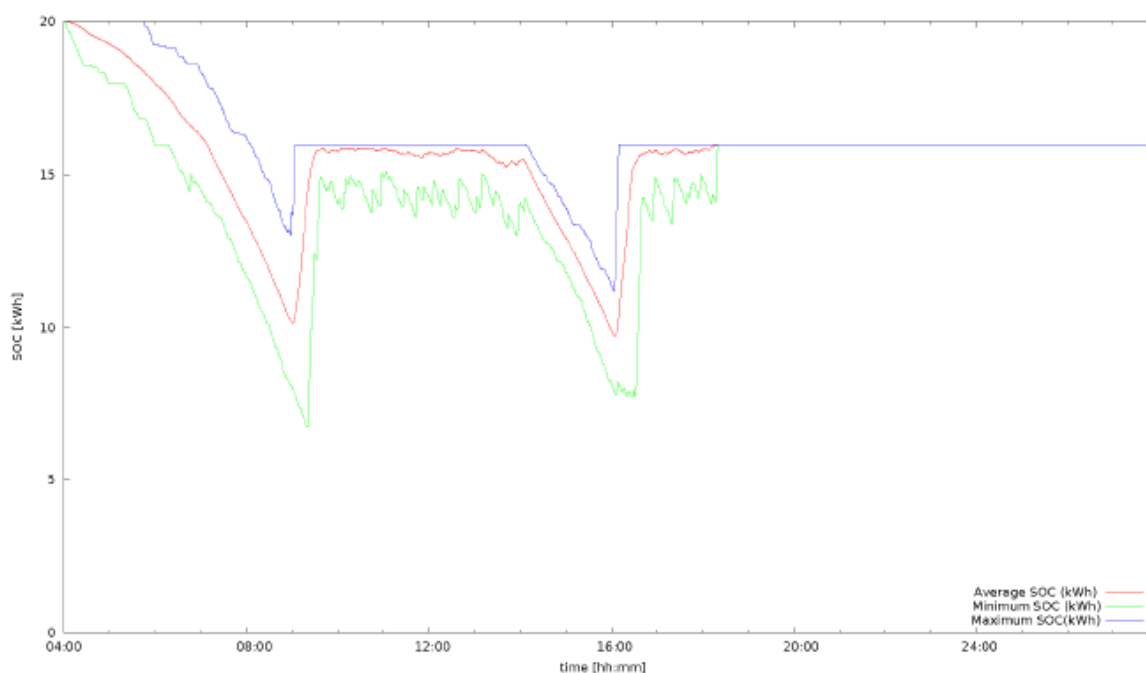


Figure 2-1 : Niveau de charge d'une flotte de taxis électriques pour une journée

Le taux de consommation d'un véhicule électrique dépend au moins de l'architecture du véhicule, du temps d'utilisation des équipements énergivores, mais aussi de l'attitude de conduite du chauffeur (Lu et al., 2012). La course est aussi mise en relation avec d'autres objets comme l'usager ou l'opérateur. La distance et durée totales et à vide des véhicules, ou encore leur vitesse moyenne sont des indicateurs de description. Leur disponibilité est étudiée : sont-ils occupés, sont-ils disponibles ? Quand les véhicules sont électriques, le nombre de véhicules en charge par heure de la journée et le niveau de charge des véhicules intéressent les chercheurs. Maciejewski (2014) modélise le niveau de charge d'une flotte de véhicules électriques sur la figure 2-1.

L'objet *borne* est décrit par de multiples indicateurs : temps d'attente à la zone de rechargement du chauffeur en fonction des types ou du nombre de bornes ou encore en fonction du niveau de demande. Mais la longueur de la file d'attente ou le taux d'occupation des sites de bornes peut aussi être sujet à évaluation.

CHAPITRE 3 MÉTHODOLOGIE GÉNÉRALE DE TRAITEMENT DE DONNÉES

Cette section est utilisée dans un premier temps pour présenter les données utilisées dans ce projet. Dans un second temps, la méthodologie de prétraitement et de traitement de données de courses et GPS sera décrite. Ainsi les différentes étapes ayant conduit à la création d'attributs seront présentées, comme l'attribut identifiant les chauffeurs dans la table GPS ou à la création d'une table GPS agrégée aux 2 min. Les enregistrements de la table GPS sont une agrégation aux 2 min de la table GPS originelle où les enregistrements sont faits toutes les 2 ou 10 secondes.

3.1 Évolution de l'offre de service de transport par taxi électrique de Téo Taxi

3.1.1 Transformation de l'offre de mobilité dans l'industrie du taxi à Montréal

Ces dernières années, l'industrie du taxi traditionnel a connu une révolution engendrée par l'utilisation des nouvelles technologies du numérique. Alors qu'auparavant les courses étaient soit héliées, soit commandées par téléphone, elles peuvent maintenant être réservées par application mobile à l'aide d'une plateforme numérique mettant en relation directe l'utilisateur et le chauffeur.

Le modèle économique d'UberX, arrivée en métropole montréalaise en octobre 2014 (RadioCanada, 2014), se base sur des chauffeurs amateurs, sans permis de taxi et travailleurs autonomes. Les clients se connectent à la plateforme numérique d'Uber et les chauffeurs prennent la course s'ils sont intéressés. Ce modèle est controversé, parce qu'il contourne la réglementation imposée à l'industrie du taxi traditionnel (The WashingtonPost, 2014). Le modèle économique de Téo Taxi, arrivé en décembre 2015, se base sur des chauffeurs salariés expérimentés. Les clients réservent leurs courses grâce à une application mobile. Téo Taxi a la spécificité supplémentaire d'être une flotte composée uniquement de véhicules électriques. Depuis août 2016, Taxi Hochelaga, Taxi Diamond et Téo Taxi sont dorénavant possédées par une seule entreprise : Taxelco.

D'autres offres peuvent entrer en compétition ou sont complémentaires à l'industrie du taxi ; ce sont des acteurs extérieurs à cette dernière. Dans les pays en voie de développement, l'offre du taxi et celle du transport en commun se concurrencent de par leur tarif similaire. Cependant dans les pays développés où le transport en commun est subventionné, le taxi peut être considéré comme un mode complémentaire au transport en commun (Darbéra, 2010). Ajoutées à ces offres de services traditionnelles, des offres innovantes ont émergé. Un nouveau modèle économique a vu le jour au Québec en 2013 (LaPresse, 2015) : les véhicules en libre-service (VLS) sur lesquels se basent les opérateurs de service Car2Go et Communauto. Le VLS se distingue de l'autopartage de la manière suivante :

- L'autopartage traditionnel permet de réserver le véhicule à l'avance et nécessite de le rapporter au même endroit (point A - point A) ; cependant, le stationnement se fait majoritairement hors rue.
- Le véhicule libre-service (VLS) permet aux usagers de faire des allers simples d'un point d'origine à différents endroits (point A - points B, C, ou D) ; il peut être stationné sur rue où la réglementation le permet, à l'intérieur du périmètre desservi. Toutefois, ce mode ne permet pas de réservation à l'avance et mise davantage sur une utilisation en courte durée par une tarification compétitive (Ville de Montréal, 2014).

Afin de réduire les émissions de gaz à effet de serre (GES), la STM et l'industrie du taxi soutiennent l'électrification du parc de VLS (Montréal, 2014). Un des buts poursuivis avec l'introduction des VLS à Montréal est la réduction du parc automobile et de l'utilisation de la voiture. Dans la région de Montréal, 27% des véhicules privés ne sont pas utilisés durant une journée ouvrable et environ 48 à 49% du parc automobile pourraient suffire à réaliser tous les déplacements de la métropole (Morency et al., 2015).

3.1.2 La compagnie de taxi électrique : Téo Taxi

Téo Taxi est une compagnie de taxis électriques lancée fin novembre 2015. Téo Taxi parie sur les véhicules électriques pour entrer sur le marché du taxi à Montréal. Son modèle économique repose sur le salariat de chauffeurs travaillant à salaire horaire fixe. Tous ses véhicules sont électriques et les modèles actuellement sur les routes sont : la Kia Soul, la Nissan Leaf et la Tesla (modèles S et X). Téo Taxi est propriétaire des véhicules et loue ou achète les permis que les chauffeurs utilisent. De plus, cette compagnie a la particularité de pouvoir assigner plusieurs véhicules à un seul permis, permettant dès lors aux chauffeurs de changer de véhicule durant leur

journée de travail. En effet, ceux-ci ont pour consigne de changer de véhicules aux bornes plutôt que d'attendre le rechargement. Par ailleurs, les courses ne pouvaient être commandées qu'à distance durant les premiers mois, mais depuis novembre-décembre 2016 le hélage est accepté.

Taxelco, fondée en 2015, détient depuis août 2016 : Taxi Hochelaga, Téo Taxi et Taxi Diamond (Téo Taxi, 2017), faisant dès lors de Taxelco un acteur majeur de la gestion de taxis de Montréal.

3.1.3 Collecte des données

La compagnie de taxi électrique Téo Taxi a débuté son implantation sur l'île de Montréal fin novembre 2015. Depuis, son réseau n'a cessé d'évoluer jusqu'au mois de décembre l'année suivante. Son territoire a connu plusieurs expansions successives comme le montre la figure 3-1, mais aussi le nombre de véhicules et de permis de taxis a augmenté comme le présenteront les statistiques descriptives de la section 3.1.4.

Les données reçues de Téo Taxi reflètent donc ce changement constant. Nous avons reçu de la part de la compagnie des données GPS et des données de courses pour fins de recherche. Les données utilisées dans ce présent mémoire se focalisent sur la période juillet-décembre 2016.

La collecte des données sur les courses d'un chauffeur, d'un véhicule et d'un usager est mise en opération à l'aide de différents appareils. Les données sont collectées par la tablette de bord du chauffeur et le taximètre. La tablette enregistre ainsi la position GPS du véhicule et les différents événements ayant lieu lors d'une journée. L'application mobile de l'usager sert à enregistrer diverses informations comme où et quand il commande une course, ou encore son numéro d'abonné. Nous ne possédons pas de données nominales sur les usagers et toutes les données sont traitées de façon strictement confidentielle.

3.1.4 Évolution de la couverture spatiale

La couverture du réseau a changé 5 fois entre le 1er décembre 2015 et le 1er janvier 2017. Chaque changement s'est opéré à l'intérieur de l'agglomération 11 de taxis. La figure 3-1 cartographie l'évolution spatiale de la couverture. La couverture initiale datant du 1er décembre 2015 contenait le centre-ville. Une première extension s'est faite le 1er janvier 2016 vers Villieray-Saint-Michel-Parc-Extension en comprenant le Plateau Mont-Royal. Une seconde extension s'est faite à partir du 1er avril, allant de Rosemont-La Petite-Patrie à Côte-des-Neiges et

vers Verdun plus au sud. À partir du 8 juin 2016, Téo Taxi a pu servir les déplacements allant de/vers l'aéroport. Enfin le dernier changement a eu lieu le 14 septembre 2016, allant de Saint-Laurent à LaSalle, au sud-ouest, à Mercier-Hochelaga-Maisonneuve, à l'est.

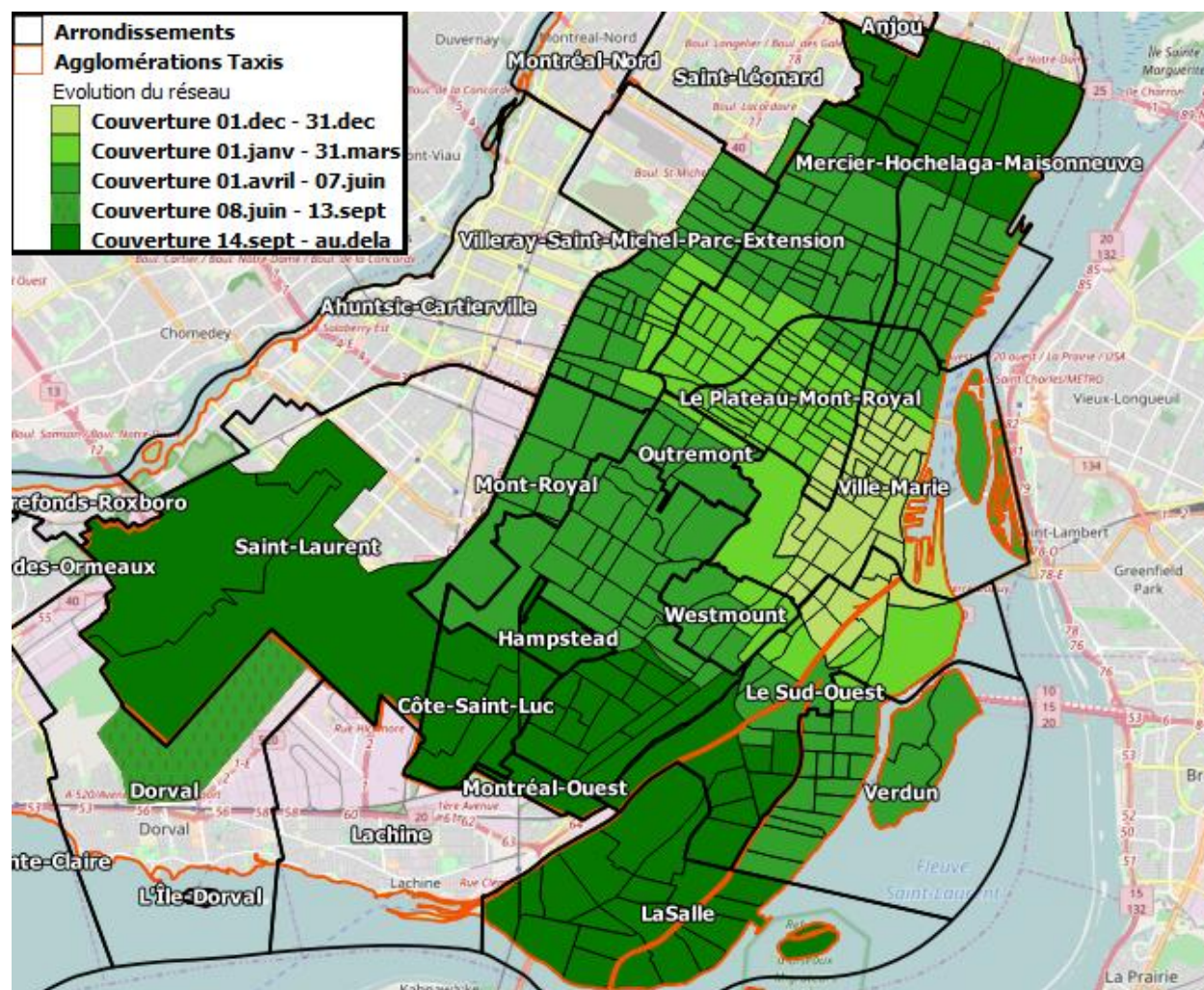


Figure 3-1 : Évolution de la couverture spatiale du réseau de Téo Taxi

3.1.5 Évolution de l'offre et de la demande

L'évolution de la couverture spatiale du réseau de transport par taxi de Téo s'accompagne de l'augmentation du nombre de chauffeurs et de véhicules au cours des mois.

Les figure 3-3 et figure 3-2 montrent cette évolution. L'augmentation du nombre total de chauffeurs (en bleu) connaît trois paliers. L'un vers fin février avec 43 chauffeurs employés. Le deuxième le 11 juillet avec 157 chauffeurs et le troisième le 11 novembre avec 278 chauffeurs

employés. En omettant l'augmentation de la couverture spatiale du 8 juin (aéroport) qui est une augmentation mineure, chaque palier semble faire suite à une extension majeure de la couverture spatiale du réseau.

Nous constatons que le nombre de chauffeurs sur paye (en rouge), c'est-à-dire le nombre de chauffeurs qui ne sont ni en congé ni absents, est égal au nombre total de chauffeurs durant les 5 premiers mois, de janvier à mai 2016, puis accuse une différence de 20-30 chauffeurs les mois suivants

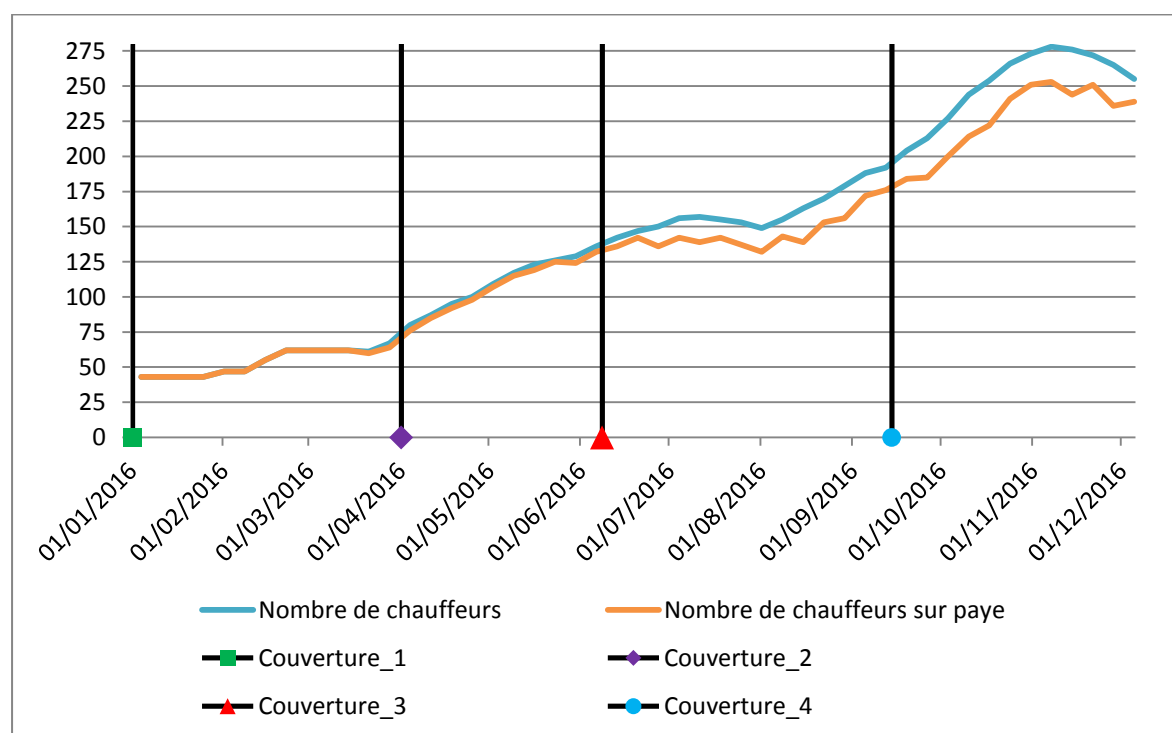


Figure 3-2 : Évolution du nombre de chauffeurs

Si le nombre total de chauffeurs et le nombre total de véhicules (en ligne pleine orange) étaient presque équivalents durant les 3 premiers mois de l'année 2016, à partir du mois de mars le nombre de chauffeurs prend son envol par rapport au nombre de véhicules. Nous remarquons d'autre part que l'augmentation du nombre total de véhicules suit l'augmentation du nombre total de chauffeurs en augmentant aussi par palier. Le nombre total de véhicules était de 60 dès début janvier. Puis fin juillet/début août, il augmente à 70 pour ensuite atteindre le 3 octobre 104 véhicules et finir en décembre avec 120 véhicules.

Le nombre total de permis suit le nombre total de véhicules tout en étant constamment un petit peu inférieur (en moyenne 10 permis de moins chaque mois par rapport au nombre de véhicules). Par exemple, quand le nombre total de véhicules était de 60 début janvier, le nombre total de permis était de 50. Et quand ce premier nombre était de 120 fin décembre, le second était de 110.

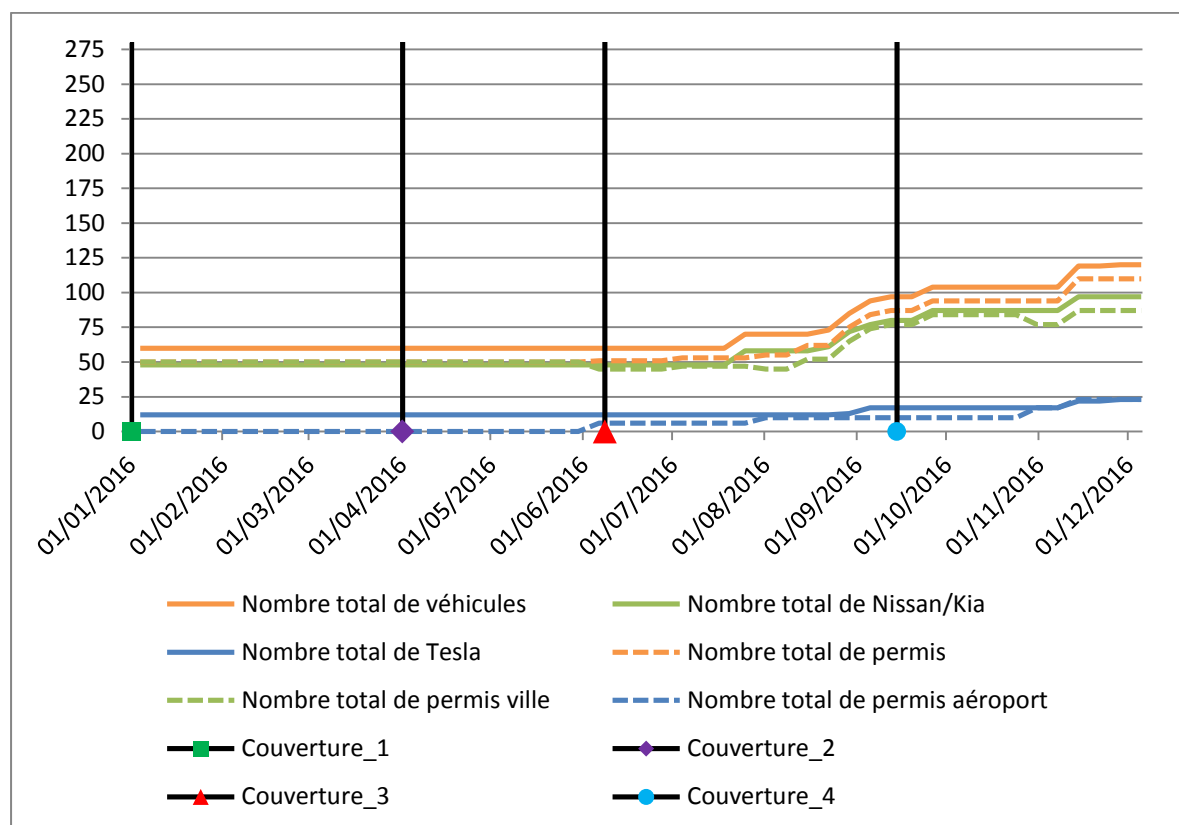


Figure 3-3 : Évolution du nombre de permis et de véhicules en fonction du temps

Pour finir, nous constatons que les évolutions du nombre de véhicules Tesla et du nombre de permis pour l'aéroport se suivent au cours des mois pour finir en décembre en nombre égal : 23 Tesla et 23 permis. Parmi les 23 Teslas, au moins 10 desservent l'aéroport d'après le communiqué fait le 8 juin 2016 par le Journal de Montréal (JournalDeMontréal, 2016).

L'évolution de l'offre, figure 3-3, s'accompagne d'une évolution de la demande figure 3-4. Comme pour l'offre, le nombre de courses réalisées augmente par palier.

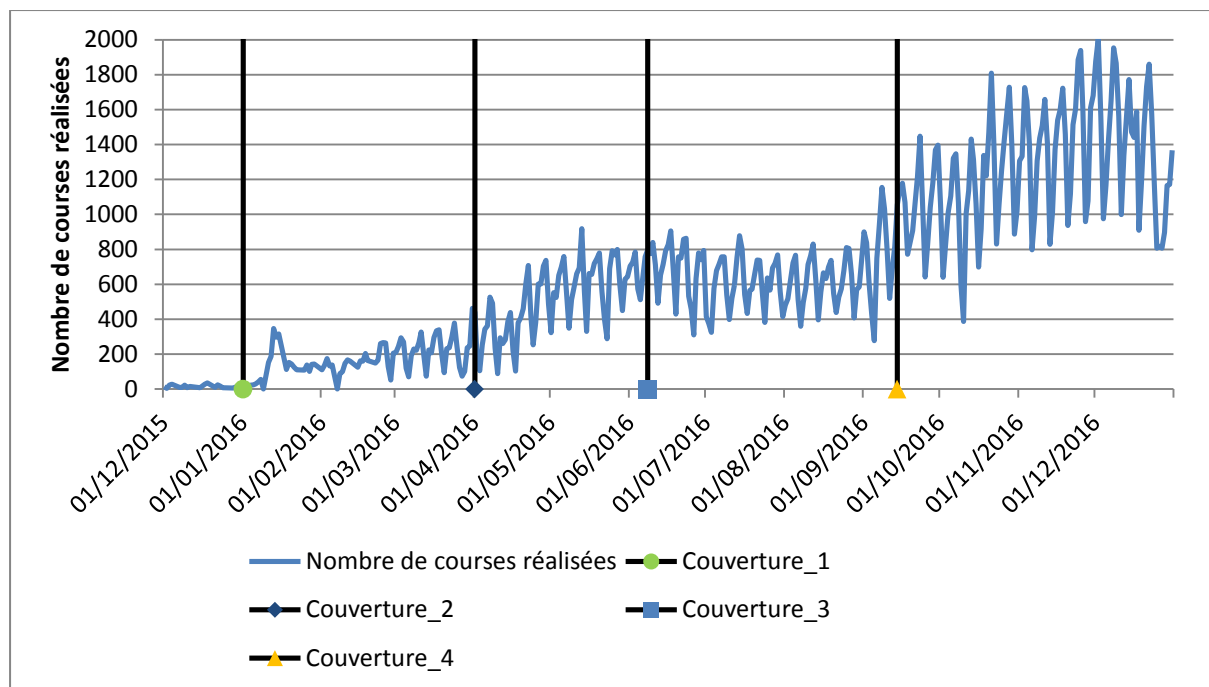


Figure 3-4 : Évolution du nombre de courses réalisées en fonction du temps

3.2 Définition d'un système de transport par taxi électrique

Cette section propose une définition des objets composant un système de taxis électriques. Elle présente ensuite les tables de courses et de données GPS qui permettent de décrire plusieurs de ces objets. Ces tables appartiennent à une base de données qui a été utilisée par la suite lors du développement méthodologique du prétraitement de données.

3.2.1 Définition des objets

Le graphique, figure 3-5, ci-dessous a été obtenu selon la méthode d'analyse orientée objet de Trépanier (1999). Cette méthode permet d'analyser chaque objet individuellement. Les objets statiques possèdent une localisation fixe dans le temps et l'espace. Les objets dynamiques sont les acteurs du transport. Les objets cinétiques sont les descripteurs du mouvement.

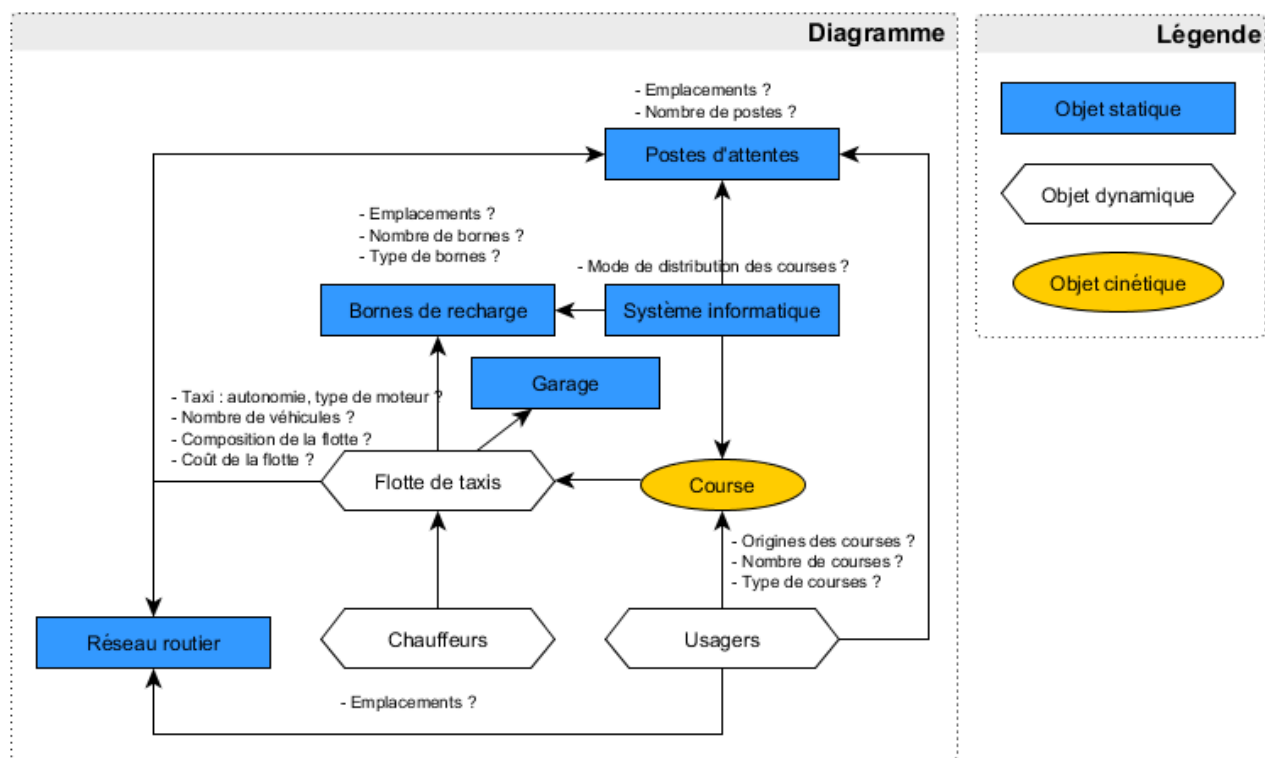


Figure 3-5 : Définition orientée-objet d'un système de transport par taxi électrique

Un réseau de taxis électriques peut être défini comme la composition de plusieurs objets. Téo Taxi est une offre de services de transport par taxi électrique. Cette offre est composée d'une *flotte de taxis* conduits par des *chauffeurs*. Les taxis électriques sont munis d'une batterie limitant

leur autonomie. Afin de pouvoir être rechargés, les véhicules sont assistés de *bornes de recharge*. Le *garage* est typiquement la zone de recharge principale ; pour ce projet c'est le garage de Saint-Patrick. Les chauffeurs sont mis en relation avec des *usagers* par l'intermédiaire d'un *système informatique*. Ce dernier est géré par un opérateur qui propose un ensemble de stratégies pour optimiser l'utilisation de la flotte. Maximiser le revenu-horaire par chauffeur peut être la ligne stratégique principale pour une compagnie privée. Cependant différentes contraintes interviennent. Chaque taxi doit toujours être suffisamment chargé pour ne pas tomber en panne ; un taxi vide doit se repositionner afin de trouver un client le plus rapidement possible, il peut donc faire le choix d'attendre à des *postes d'attente*.

Le *réseau routier* permet de faire le lien physique entre l'offre et la demande. L'ensemble des usagers forme la demande. La demande émet des requêtes pour effectuer des *courses*. Ces courses possèdent notamment une origine, une destination et une heure de commande.

3.2.2 Description de la base de données

La base de données étudiée dans ce projet est composée principalement d'une table de courses et d'une table GPS. Celles-ci ont chacune des attributs décrivant leur objet. L'objet décrit par la table de course est la *course* d'un *usager* satisfaite par un *chauffeur* à l'aide d'un *véhicule*. Ainsi nous possédons des informations sur quatre objets. L'objet principal d'une table GPS est le *point GPS*. Ce point décrit l'emplacement d'un *véhicule* à une heure donnée. Dans cette table, aucune information sur l'identifiant numérique du chauffeur n'est disponible.

3.2.2.1 La table des courses

Les données de courses sont séparées en 2 fichiers distincts : l'un comportant les données des courses corporatives réalisées pour le service d'une entreprise et l'autre comportant les données des courses au service des particuliers. Ces deux tables de données distinctes ont été agrégées en une seule table dont les attributs qui la composent sont présentés dans le tableau 3-1 ci-dessous.

Tableau 3-1 : Informations données par la table des courses

Informations sur les courses	Caractéristiques
ID	Identifiant de la course
Départ	Coordonnées et Horodatage
Arrivée	Coordonnées et Horodatage
Usager	Identifiant, Type d'utilisateur, Temps d'attente et Évaluation de la course
Caractéristiques de la course	Type de service, Type de course, Zone couverte, Horodatage et Durée et coût de la course
Annulation	Horodatage de l'annulation et Raison d'annulation
Chauffeur	Identifiant et Horodatage de certaines actions
Véhicule	Identifiant du véhicule

Chaque course a un identifiant. Le départ et l'arrivée de celle-ci sont caractérisés par une adresse et un horodatage. À chaque course est associé un identifiant d'utilisateur. Nous savons de plus si l'utilisateur est un particulier ou le représentant d'une entreprise, quel a été son temps d'attente avant l'arrivée du taxi et la qualité de la course perçue par l'utilisateur (de 1 à 5 étoiles). Quand un utilisateur commande une course, plusieurs attributs sont pris en compte. Le type de services rendu indique si la course est une course aéroportuaire ou régulière. Les courses aéroportuaires sont celles qui vont ou qui proviennent de l'aéroport. Les courses régulières sont celles qui n'ont pas d'origine ou de destination fixée au préalable par Téo Taxi. La table de données indique aussi si la course est commandée par *hélage* ou à distance. Une course peut être commandée à distance par *appel téléphonique*, par *application mobile*, ou encore par le biais du *site internet de Téo Taxi* pour certaines entreprises. Nous savons aussi quand la course a été commandée, combien de temps elle a duré, son coût et quelle a été la taille du territoire couvert par le réseau de Téo Taxi lorsque la course a été faite. De plus il est précisé si les commandes ont été réalisées ou annulées. Dans le cas des courses annulées : l'horodatage de l'annulation, le temps attendu avant annulation et la raison de l'annulation sont donnés.

Enfin, les identifiants du chauffeur et de son véhicule sont présents. Des horodatages précisent les actions du chauffeur : quand il a accepté la commande, quand il est arrivé à l'origine de la course, ainsi que le temps passé à atteindre le point d'origine de son client.

3.2.2.2 La table GPS

La table GPS est composée de peu d'attributs, mais comporte des données massives. Chaque point GPS trace un véhicule dans le temps et dans l'espace. Or près de 100 véhicules électriques par jour sont en service sur la route et le dispositif GPS enregistre la position de chacun d'eux toutes les 2 secondes en moyenne quand ils sont en mouvement, toutes les 10 secondes en moyenne quand ils sont à l'arrêt.

Tableau 3-2 : Informations données par la table GPS

Informations GPS	Caractéristiques
ID	Identifiant du point GPS
Position (espace et temps)	Horodatage et Position (lat, lon) du véhicule
Véhicule	Identifiant du véhicule
Statut	Statut de disponibilité du véhicule

Chaque point GPS décrit la position d'un véhicule dans le temps et dans l'espace. Chaque point possède donc un identifiant unique. Il lui est attribué l'heure et la date à laquelle il a été enregistré dans la table, sa position géographique (latitude, longitude) ainsi que l'identifiant du véhicule qu'il trace.

3.2.2.3 Structure de la base de données relationnelle

La figure 3-6 ci-dessous présente la base de données relationnelle reliant l'ensemble de ces différentes tables. Aux tables transmises par Téo Taxi s'ajoutent les tables suivantes : la table des secteurs de recensement, des secteurs municipaux, des arrondissements, des postes d'attente, des zones de recharge et du calendrier. Chacune est la table d'un objet circonscrit à l'aide de plusieurs attributs. Les relations intertables sont soit de type 1-1, soit de type 1-plusieurs ou soit de type plusieurs à plusieurs.

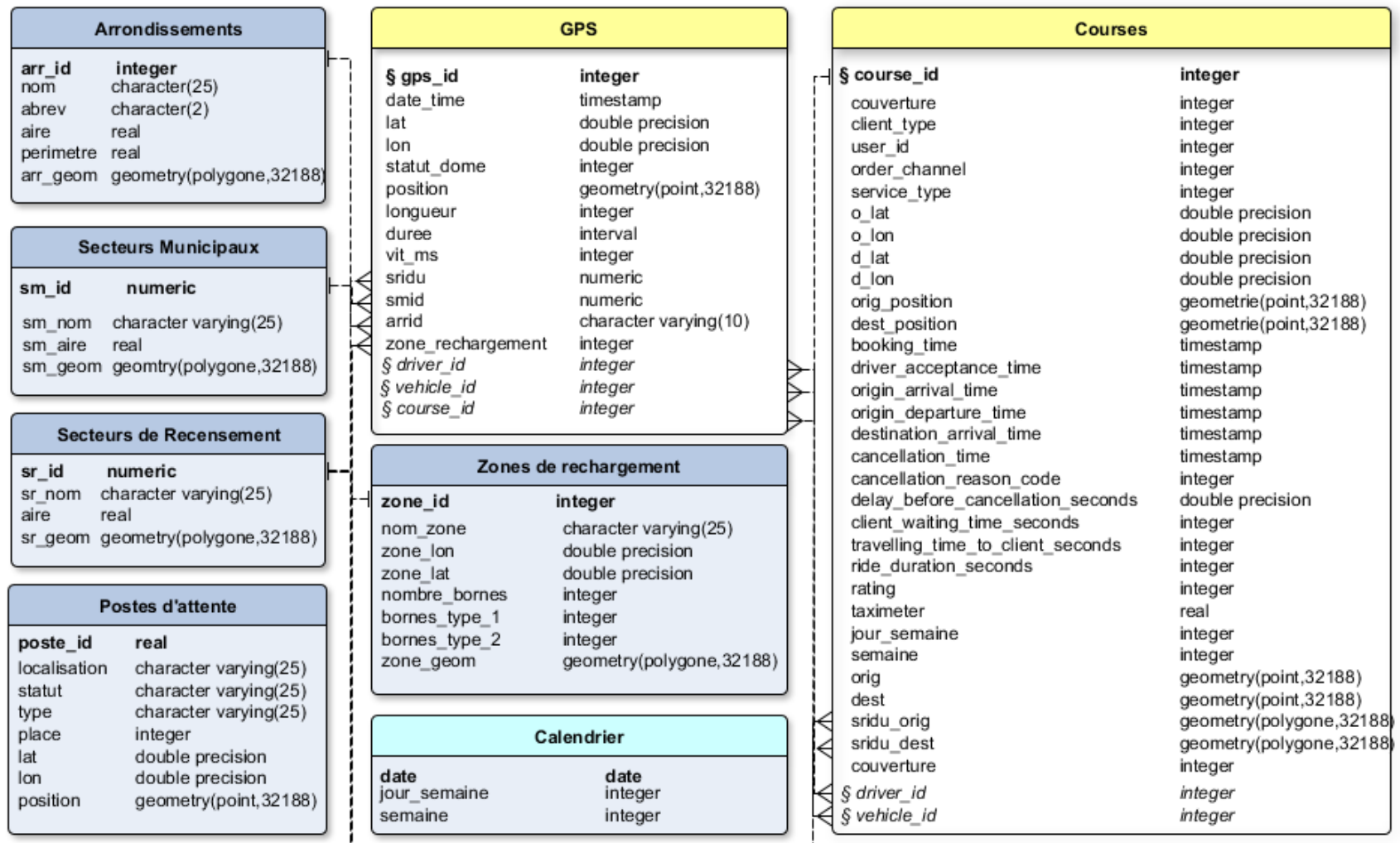


Figure 3-6 : Diagramme de la base de données relationnelle de taxis électriques

3.3 Méthodologie générale de prétraitement de données

La première étape, lorsqu'une base de données est fournie, est de la traiter. La figure 3-7 ci-dessous schématise la méthodologie employée pour traiter les données GPS et de courses en taxi électrique. Dans le cadre des données reçues pour ce projet, 4 étapes de prétraitement de données se succèdent.

Étape 1 : La première étape consiste à traiter la table de données GPS. Les enregistrements sont importés dans un logiciel de gestion de base de données - ici PostgreSQL - pour être ensuite nettoyés et standardisés. Ces sous-étapes permettent de s'assurer d'avoir une table propre, c'est-à-dire sans doublon et où les enregistrements incomplets sont repérés. Un enregistrement incomplet est une ligne de données qui n'a pas de valeur pour un ou plusieurs attributs alors que cela serait pourtant attendu. À la fin de cette première étape, comme à la fin de toutes les étapes suivantes, la suppression des tables temporaires est requise pour ne pas s'encombrer de tables inutiles dans la base de données. Le nombre d'enregistrements était de 231 066 869 avant prétraitement, 0,3% des enregistrements ont été supprimés après prétraitement.

Étape 2 : La deuxième étape est le traitement de données des courses en taxis électriques. La même procédure que pour l'étape précédente est appliquée : une importation des données, un nettoyage, une standardisation et une suppression des tables temporaires. Lors de la standardisation, un attribut *longueur* qui donne la longueur de chaque course est construit à partir de la table GPS. Pour chaque véhicule, cette longueur est définie comme la somme des distances qui séparent chaque point GPS qui compose le trajet de sa course. Ici, le nombre d'enregistrements était de 262 484 avant prétraitement, 0,9% des enregistrements ont été supprimés après prétraitement.

Étape 3 : La table GPS précédemment traitée n'a pas d'attribut identifiant les chauffeurs de véhicules (*driver_id*), c'est-à-dire qu'aucun chauffeur n'est attribué au véhicule tracé par le dispositif GPS. Il est cependant possible de reconstruire cet attribut à partir de la table course. C'est pourquoi cette étape, qui finit de standardiser la table GPS, vient après l'étape du prétraitement des courses. La méthodologie de reconstruction de l'attribut identifiant les chauffeurs sera expliquée dans la section 3.5.

Étape 4 : Enfin, la quatrième et dernière étape de prétraitement de données consiste à réduire la taille de la table GPS. En effet le calcul des différentes requêtes SQL est ralenti par la quantité d'enregistrements qu'elle contient. Les enregistrements GPS sont aux 2 secondes quand le véhicule est en mouvement et aux 10 secondes quand le véhicule est à l'arrêt. Une table GPS qui agrège les enregistrements sur 2 min permet de considérablement diminuer sa taille - de plus de 12 fois sa taille si tous les enregistrements étaient effectués aux 10 secondes. Le nombre d'enregistrements était donc de 230 395 852 avant prétraitement, 96 % des enregistrements ont été agrégés après prétraitement, faisant passer le nombre d'enregistrements à 8 580 002. La section 3.6 présentera plus amplement la méthodologie de construction de cette table. Ce niveau d'agrégation n'affectera pas les estimations faites par la suite, car les objets étudiés ne dépendent pas du nombre d'enregistrements.

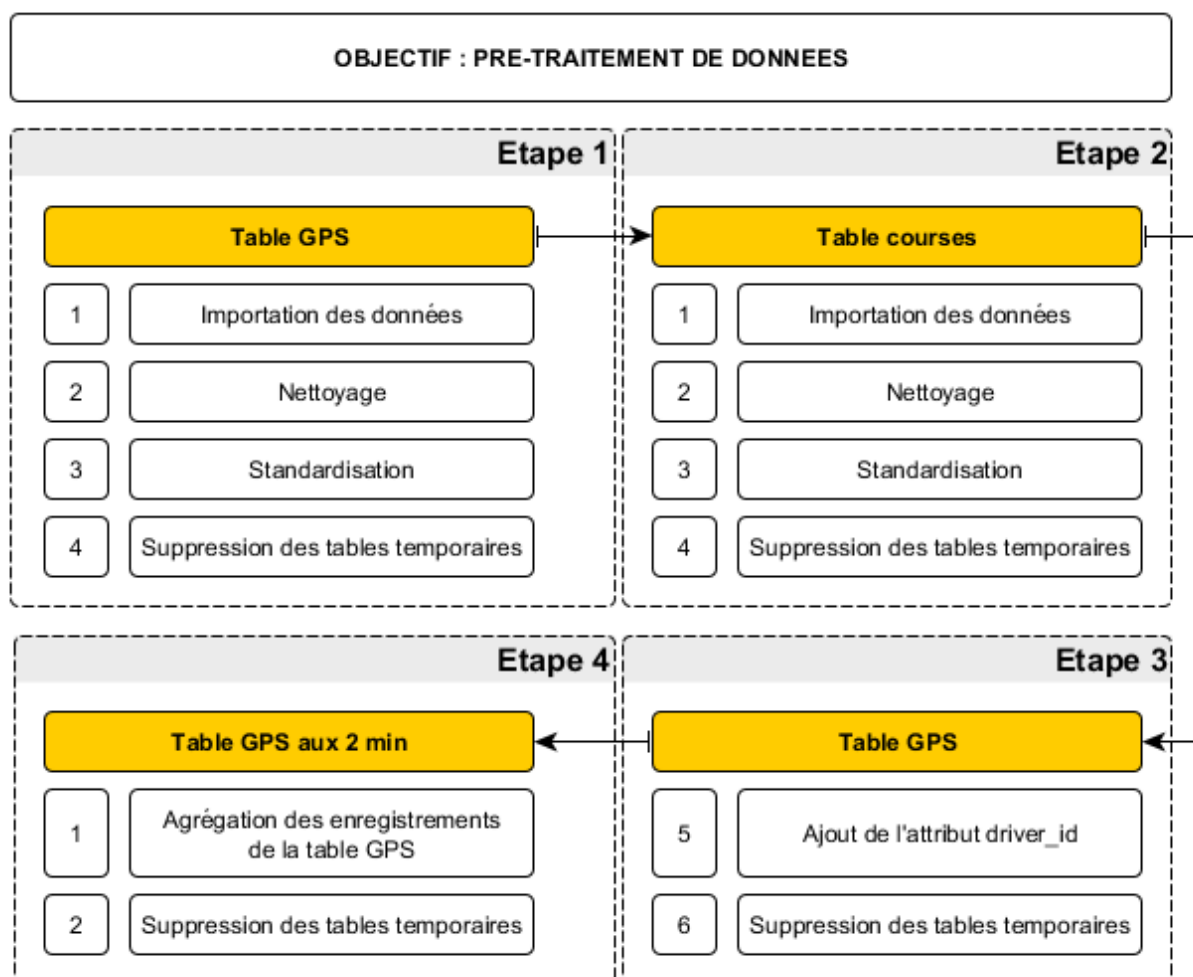


Figure 3-7 : Méthodologie de prétraitement des données GPS et de courses de taxis

À la suite de cette procédure de prétraitement de données, la base de données est prête à être utilisée pour le calcul de différentes requêtes.

3.4 Standardisation et nettoyages des données

Avant de pouvoir utiliser la table des courses de Téo Taxi pour l'analyser, cette dernière doit subir un prétraitement afin de la préparer aux requêtes SQL qui lui seront appliquées. Une standardisation et un nettoyage de données sont dans ce cas nécessaires. Il est entendu par standardisation des données, la modification d'une table pour la rendre plus adéquate au travail d'analyse. Cette modification n'engendre pas de perte de données. Le nettoyage de données supprime les données qui ont parfois une valeur aberrante due à des erreurs de dysfonctionnement du système automatisé de collecte de données, mais aussi souvent dû aux erreurs humaines. Par exemple, parfois, l'usager oublie d'annuler sa course et le chauffeur qui aurait dû la réaliser ne l'annule pas non plus. Nous obtenons alors pour cette course un temps d'attente avant annulation qui peut durer plus d'une heure.

3.4.1 Les règles de standardisation des données

Les règles de standardisation pour la table des courses sont les suivantes :

- Agréger la table des courses corporatives avec la table des courses pour particuliers. Créer un attribut *client_type* pour définir si la course est corporative ou pour un particulier.
- Créer un attribut *couverture* afin d'indexer chaque course à une taille de couverture du réseau. Rappelons que depuis le début de l'opérationnalisation de Téo Taxi, la couverture du réseau a évolué avec le temps.
- Harmoniser l'ordre des attributs de la table de courses corporatives avec la table de courses des particuliers.
- Convertir dans le système EPSG 32188 (NAD83 MTM Zone 8) toutes les coordonnées de localisation qui sont dans le système de géoréférencement EPSG 4326 (longitude, latitude), ou tout autre système de coordonnées. Ceci permettra d'éviter les distorsions liées à l'étude de points venant de tables ayant des systèmes de géoréférencement différents.
- Suppression possible des attributs : *client_waiting_time*, *ride_duration*, *delay_before_cancellation_seconds* et *travelling_time_to_client_seconds*. Les valeurs de ces champs sont issues d'opérations sur d'autres champs de cette même table.

- Transformation possible des valeurs nominales de l'attribut *order_channel* en valeurs numériques.

3.4.2 Nettoyage des données : les règles d'exclusion

3.4.2.1 Nettoyage de la table GPS

Il est entendu par *règle d'exclusion*, une règle qui sert à nettoyer une table de données. La table GPS est soumise aux trois règles d'exclusion suivantes :

- Suppression des doublons ;
- Suppression des points n'ayant pas de longitude ou de latitude (ou les deux) ;
- Suppression des points ayant une longitude ou latitude = 0.

Après nettoyage, la table des courses passe de 262 484 enregistrements à 260 250. 0,9% des enregistrements a été supprimé.

3.4.2.2 Nettoyage de la table de courses

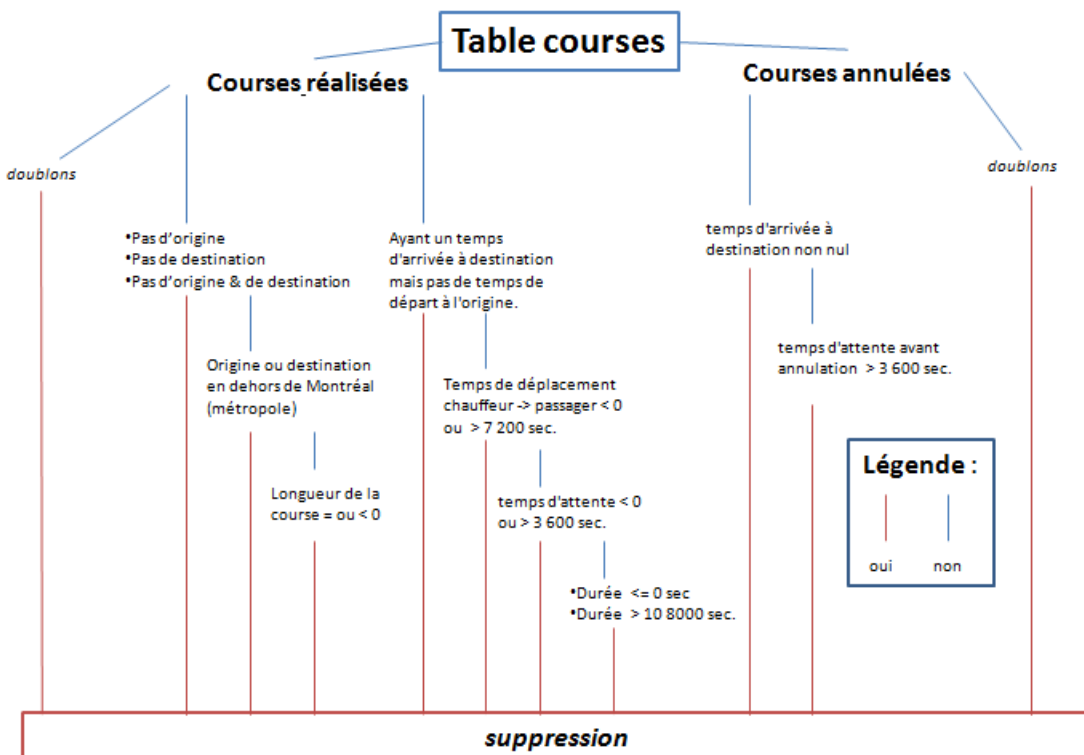


Figure 3-8 : Graphique résumant les règles de nettoyage de la table des courses

La figure 3-8 présente les règles d'exclusion soumises à la table des courses. Les règles d'exclusion sont basées sur les distances et temps aberrants. La première d'exclusion est la suppression des doublons. Les autres peuvent se faire sur la base de la distance, du temps et de l'annulation et la réalisation des courses.

Suppression sur la base de la distance :

- Suppression des courses ayant des origines et destinations vides.
- Suppression des courses dont la distance = 0 ou < 0 km.
- Suppression des courses hors du territoire d'étude. Cette étape permet en plus de supprimer les points de localisations aberrants. Certains peuvent par exemple se trouver en plein milieu de l'Océan Pacifique.

Suppression sur la base du temps :

- Suppression des courses réalisées, avec temps d'arrivée à destination, mais sans temps de départ à l'origine.
- Suppression des courses dont la durée ≤ 0 s et $> 10\,800$ s. Cette dernière valeur maximale arbitraire est posée par Téo Taxi. Elle correspond à la valeur qui devrait couvrir toutes les courses régulières. Quand la durée dépasse cette valeur, Téo Taxi met une valeur vide. Il s'agit donc aussi de supprimer ces courses à valeur vide.
- Suppression des courses dont le chauffeur a mis un temps de déplacement jusqu'à son passager qui n'est pas entre 0 à $7\,200$ s. Valeur maximale arbitraire posée par Téo Taxi.
- Suppression des courses dont le client a un temps d'attente qui n'est pas dans l'intervalle de 0 à $3\,600$ s. Valeur maximale arbitraire posée par Téo Taxi.

Nous pouvons aussi créer des règles d'exclusion de courses sur la base du temps pour les courses annulées.

Suppression sur la base des courses annulées :

- Suppression des courses annulées qui semblent être des courses réalisées. C'est-à-dire toutes les courses qui, bien qu'elles soient annulées, possèdent un temps d'arrivée à destination.

- Possibilité de supprimer les courses dont le temps d'attente avant annulation est $> 3\,600$ s. En effet, Téo Taxi pose que le temps d'attente de l'usager doit être dans l'intervalle de 0 à $3\,600$ s pour n'importe quelle course. En particulier, si un usager met plus de $3\,600$ s avant d'annuler sa course, nous considérons cette dernière comme erronée.

A la suite de ce nettoyage, la table GPS est passée de 231 066 869 à 230 395 852 enregistrements. 0,3% des enregistrements a été supprimé.

3.5 Table GPS : ajout de l'attribut identifiant le chauffeur du véhicule

Cette section a pour but de répondre à un problème opérationnel dans le traitement de données. En effet, les tables GPS fournies par Téo Taxi ne contenaient pas l'identifiant du chauffeur du véhicule. Il était possible d'identifier le véhicule, mais pas le chauffeur. À partir de ce constat, une solution a été établie pour retrouver l'identifiant du chauffeur conduisant le véhicule pour chaque ligne GPS.

3.5.1 Comment reconstituer l'attribut identifiant le chauffeur d'une table GPS ?

Dans la politique de la compagnie Téo Taxi, un véhicule n'est pas associé à un chauffeur unique. Au contraire un véhicule peut avoir plusieurs chauffeurs au cours d'une même journée. En effet lorsque la réserve d'énergie du taxi est basse, il retourne à une borne de recharge où il sera rechargé. Pour une question d'efficacité opérationnelle et économique, le chauffeur n'attend pas que son véhicule se recharge. Au lieu de recharge, il lui est proposé un second véhicule à batterie pleine ou suffisamment rechargé avec lequel il va continuer son quart de travail. Par conséquent à partir de cette information, nous pouvons supposer qu'à chaque fois qu'un véhicule *sort* d'une zone de recharge il change de chauffeur.

Si une borne de recharge est localisée par un point, une zone de recharge est localisée par une surface. La zone de recharge est une zone a priori couverte par les taxis ayant besoin de se recharger à une borne. Ne connaissant pas la date et l'instant à partir desquels un véhicule est connecté à une borne, nous ne pouvons pas déterminer l'instant où le chauffeur est absent du véhicule. C'est pourquoi pour pallier ce manque d'information temporelle, il est possible d'utiliser

l'information spatiale. En effet, connaissant les lieux où se trouvent les bornes de recharge, nous pouvons définir des zones de recharge. Et comme montré dans le paragraphe ci-dessus, nous pouvons alors définir l'instant où un véhicule a changé de chauffeur comme étant l'instant où le véhicule sort de la zone de recharge.

Quel *identifiant chauffeur* sera associé au véhicule lorsqu'il sortira de la zone ? Nous savons que si un chauffeur sort d'une zone de recharge alors il est possible que des demandes de courses lui soient faites. Or dans la table de courses, à chaque course sont associés les numéros du véhicule et du chauffeur l'ayant réalisée. Par conséquent lorsque dans la table GPS nous repérons l'instant où le véhicule est sorti de la zone, nous pouvons dès lors lui associer *l'identifiant chauffeur* des courses qui seront faites par la suite avec ce même véhicule. La figure 3-9 illustre le raisonnement.

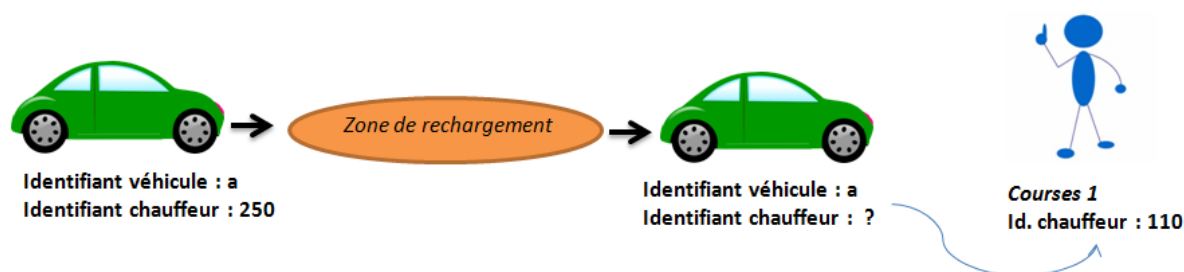


Figure 3-9 : Méthodologie de récupération de l'identifiant chauffeur

3.5.2 Calibration des zones de recharge

Les sites de rechargement sont donnés par la compagnie Téo Taxi. Cependant, comment définir le contour de ces zones ? Il est nécessaire de convenir d'une zone qui ne soit pas trop grande pour éviter le passage éphémère de véhicules, mais aussi d'une zone qui ne soit pas trop petite pour pouvoir récupérer les points GPS de tous les véhicules à l'arrêt.

Il y a deux façons de décider de la taille et de la forme de la zone. L'attribut *statut_dome* permet d'identifier géographiquement une zone convenable de rechargement. Lorsque le *statut_dome* = 7, le chauffeur se connecte au véhicule. Les points GPS associés à ce statut indiquent donc les positions où le taxi s'est arrêté pour se recharger. Afin de valider le contour des zones, il est aussi possible d'utiliser des tables de connections aux bornes de recharges. Téo Taxi dispose de tables de connexions aux bornes par véhicule. La structure des données la rend difficilement exploitable

à grande échelle, mais elle reste utilisable pour de petites analyses. En l'occurrence, en observant pour une dizaine de véhicules les lieux où ils se trouvaient lorsqu'ils étaient en connexion, il est possible de valider les contours de zones préalablement choisies.



Figure 3-10 : Vue de la zone de recharge de l'aéroport. Vue cartographiée contre vue satellite.

La figure 3-10 ci-dessus présente un exemple de zone de recharge. Nous savons que la borne de recharge de l'aéroport se trouve à l'intérieur d'un stationnement (à droite). La zone de recharge associée est la surface verte (à gauche).

3.5.3 Méthodologie de reconstitution de l'attribut identifiant le chauffeur du véhicule

Supposons que l'attribut identifiant le chauffeur se nomme *driver_id*. La figure 3-11 présente la méthodologie utilisée pour reconstituer l'attribut *driver_id* dans la table GPS. Elle se présente en 3 étapes.

La première étape consiste à créer 2 nouvelles tables GPS. En effet pour chaque jour de données, la reconstitution du *driver_id* a besoin du jour précédent et du jour suivant. Ainsi, selon la période temporelle en traitement, il est nécessaire d'avoir le jour précédent et le jour suivant cette période. Pour la table GPS de la période temporelle analysée et pour les tables GPS des jours précédents et suivants, l'étape 1 consiste à ajouter des attributs caractérisant la zone de recharge avec un attribut que l'on peut nommer *zone_rechargement*.

L'attribut *zone_rechargement* vient indiquer si le véhicule est dans une zone de recharge ou non et si oui, laquelle. Dans l'optique de reconstituer l'attribut *driver_id*, nous devons repérer les instants où les véhicules sortent des zones. Pour se faire, nous construisons l'attribut *code_zone* indiquant 1 si le véhicule est dans la zone et 0 sinon. Puis nous créons l'attribut *code_changement*

qui, en fonction de l'attribut *code_zone*, indiquera 1 si le véhicule entre dans une zone, -1 s'il sort et 0 s'il est soit à l'intérieur soit à l'extérieur. Ayant ces dernières informations nous savons dorénavant quand chaque véhicule entre et sort d'une zone au cours d'une journée. Il suffit alors d'ajouter le *driver_id* adéquat pour toute la durée où le véhicule est à l'extérieur d'une zone, c'est-à-dire pour tous les enregistrements GPS entre l'instant où *code_changement* = -1 et *code_changement* = 1.

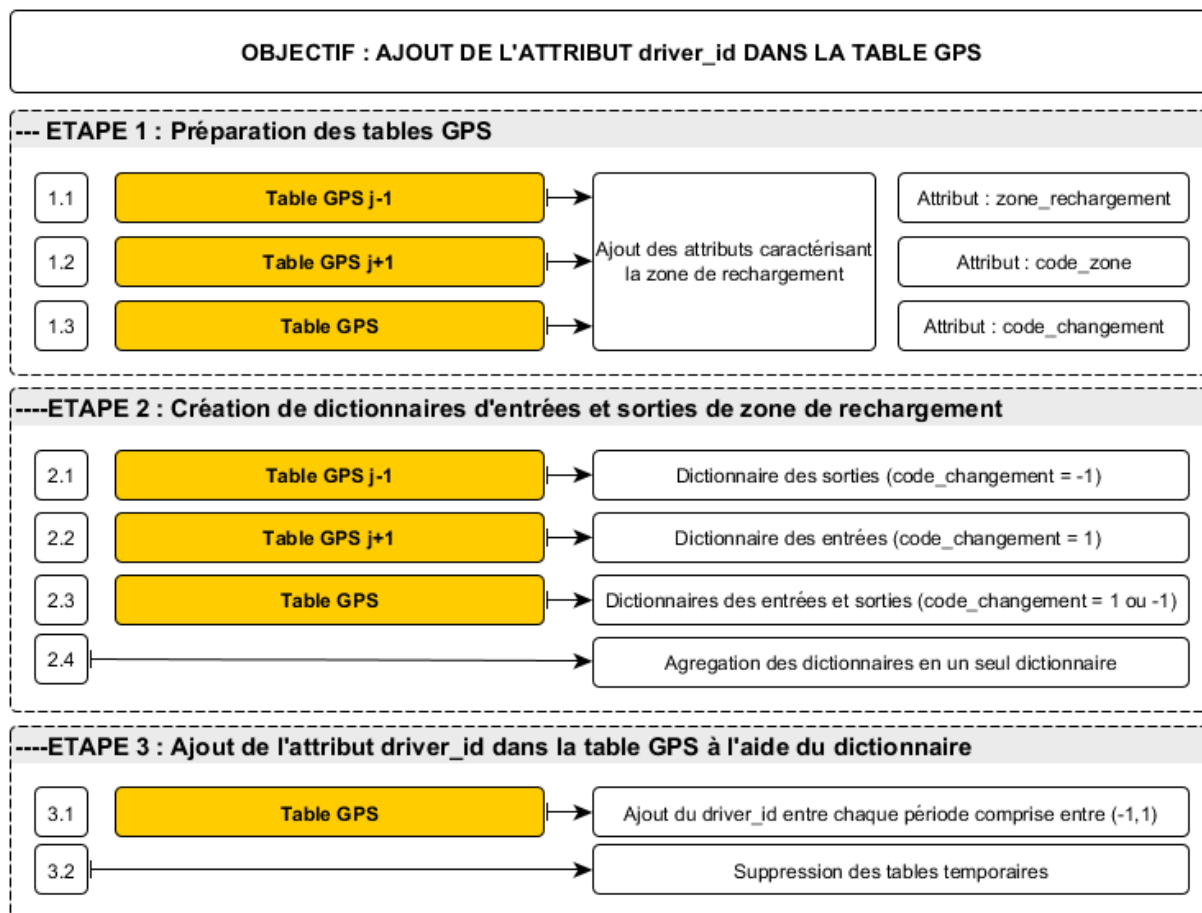


Figure 3-11: Méthodologie permettant de reconstituer l'attribut *driver_id* dans la table GPS

La figure 3-12 présente un extrait d'une table GPS où le *driver_id* a été reconstitué.

		temps timestamp without time zone	vehicle_id integer	statut_dome integer	zone_rechargement integer	code_zone integer	code_changement integer	driver_id integer
416574		2016-10-01 15:35:12		1	1	1	0	0
416575		2016-10-01 15:35:14	a	1	1	1	0	0
416576		2016-10-01 15:35:16	a	1	1	1	0	0
416577		2016-10-01 15:35:18	a	1		0	-1	b
416578		2016-10-01 15:35:20	a	1		0	0	b
416579		2016-10-01 15:35:22	a	1		0	0	b

Figure 3-12 : Extrait d'une table GPS reconstituée

Afin d'automatiser ce procédé, un dictionnaire de sorties et d'entrées des véhicules est créé. C'est l'étape 2. Enfin l'étape 3 consiste à utiliser ce dictionnaire pour attribuer à tous les véhicules le *driver_id* qui lui est associé, en fonction des courses qui ont été faites pendant la période où chaque véhicule est en dehors des zones de recharge.

3.5.4 Limite de la méthode

Cette méthode de reconstitution comporte néanmoins quelques limites à noter. Il n'est pas possible de retrouver les identifiants des chauffeurs dont le véhicule n'a pas commencé sa journée dans une zone de recharge. Ce peut être le cas des nouveaux véhicules venant s'ajouter à la flotte régulière de Téo Taxi. Il faut alors attendre qu'il sorte d'une zone de recharge pour pouvoir recommencer à déterminer les *driver_id* qui lui sont associés.

La méthode manque aussi de précision étant donné le calibrage des zones de recharge. Rappelons qu'une zone de recharge doit être dessinée de telle manière qu'elle ne soit pas trop grande pour éviter le passage éphémère de véhicules, mais ne doit aussi pas être trop petite pour pouvoir récupérer les points GPS de tous les véhicules à l'arrêt.

Par ailleurs, il n'est pas possible de reconnaître le chauffeur des véhicules sortant d'une zone et n'ayant pas eu de courses par la suite. Les identifiants de chauffeurs étant liés aux courses, sans course il n'est pas possible de reconstituer le *driver_id*.

Enfin, chaque fois qu'un véhicule est à l'intérieur d'une zone, l'attribut *driver_id* est arbitrairement égal à 0 parce qu'il n'est pas possible de déterminer le *driver_id* à l'intérieur des zones avec la méthode proposée.

Si nous disposions d'une table de connexion aux bornes de recharge, il serait alors possible de reconstituer avec plus de précision l'attribut *driver_id* à partir des temps de connexion aux bornes de recharge.

3.6 Construction de la table GPS agrégée

3.6.1 Justification de la création d'une table GPS agrégée aux 2 min

La construction de la table GPS agrégée permet de réduire la taille de la table GPS de base. Cette table est donc une agrégation des données aux 2 minutes. Cette section présente les attributs que contient cette table.

Chaque enregistrement de cette table regroupe les enregistrements des 2 minutes à venir. C'est-à-dire qu'en lisant un enregistrement au temps t de la table aux 2 min, nous obtenons des informations sur ce qu'il s'est passé du temps t au temps $t + 2$ min de la table GPS de base. Le but de la création de cette table est de gagner en temps de calcul durant les analyses faites sur les données GPS.

Pourquoi avoir choisi une agrégation aux 2 min plutôt qu'au 1 min ou aux 5 min ? La décision du niveau d'agrégation est arbitraire. Cependant, le but étant de gagner en temps de calcul et de rendre pratique l'utilisation de la table, une agrégation au 1 min paraissait insuffisante. De l'autre côté, une agrégation aux 5 min aurait pu être faite afin de réduire encore plus la taille de la table. Finalement une agrégation aux 2 min permettait de gagner en taille d'agrégation tout en réduisant les approximations liées à la pondération. La pondération est expliquée dans la section suivante.

3.6.2 Présentation des attributs

La figure 3-13 présente le nom, le type et la description de l'ensemble des attributs de la table GPS aux 2 min. Celle-ci contient 36 attributs. Certains attributs peuvent se retrouver à partir d'autres attributs (comme l'attribut *date* à partir de l'attribut *temps*). Cependant, ils sont conservés, car ils sont utiles ultérieurement pour gagner du temps dans l'obtention d'analyses.

Pondération

De manière générale, dans un intervalle de 2 min, chaque fois qu'un attribut pouvait prendre plusieurs valeurs, la valeur qui a affecté le plus de points GPS a été sélectionnée pour être affichée dans la table GPS aux 2 min. Par exemple, la valeur du *statut_dome* ayant affecté le plus de points GPS durant un intervalle de 2 min sera sélectionnée. Les valeurs ayant affecté le moins de points seront ignorées. Par conséquent, avec cette manière de pondérée, une approximation est faite.

Cette pondération a été appliquée aux attributs *statut_dome*, *driver_id*, *sridu*, *smid* et *arrid* qui respectivement identifient le statut de disponibilité du véhicule, le chauffeur du véhicule, le secteur de recensement, le secteur municipal et l'arrondissement traversés.

Vitesse

L'agrégation des vitesses sur 2 min nous invite à créer un attribut *vit_moy* indiquant la vitesse moyenne (moyenne des vitesses instantanées) à laquelle le véhicule roulera durant les 2 prochaines minutes. Les vitesses au 25e, 50e et 75e centile donnent des indications sur la vitesse du véhicule dans chaque groupe de vitesses agrégées.

Durée

L'attribut *durée* indique la durée passée par le chauffeur à conduire le véhicule durant la période couvrant les 2 min d'agrégation. Il est à noter que pour chaque véhicule, la durée séparant deux traces GPS est soit d'environ 2 s, soit d'environ 10 s. Cependant, parce qu'il y a des périodes de déconnexion, deux traces GPS peuvent être séparées par plusieurs minutes, voire plusieurs heures. Ainsi, pour éliminer les temps de pause, il est nécessaire de ne pas prendre en compte les durées supérieures à une certaine valeur, ici, 30 s.

Table GPS au 2 min		
§ gps_id	integer	Identifiant de l'enregistrement GPS
vehicule_id	integer	Identifiant du vehicule tracé.
annee	integer	Année d'enregistrement.
mois	integer	Mois d'enregistrement.
jour	integer	Jour d'enregistrement.
date	date	Date ('a-m-j') d'enregistrement.
hr	integer	Heure d'enregistrement.
min	integer	Minute d'enregistrement.
temps	timestamp	Temps ('a-m-j h:m:s') d'enregistrement.
sem	double precision	Identifiant de la semaine de l'année.
jr_sem	double precision	Identifiant du jour de la semaine.
driver_id	integer	Identifiant du conducteur de véhicule.
vit_moy	bigint	Vitesse moyenne du véhicule (m/s).
vit_25c	double precision	25e centile de la vitesse du véhicule.
vit_50c	double precision	50e centile de la vitesse du véhicule.
vit_75c	double precision	75e centile de la vitesse du véhicule.
duree	interval	Durée passée à conduire couvrant la période de 2 min.
nb_points_gps	bigint	Nombre d'enregistrements couvrant la période de 2 min.
points_2min	geometry	Points GPS agrégés au 2 min.
ligne_2min	geometry	Ligne GPS agrégée au 2 min.
distance	integer	Distance passée à conduire couvrant la période de 2 min.
statut_dome	integer	Statut de disponibilité prépondérant du véhicule.
nb_points_gps_0	bigint	Points GPS agrégés au 2 min (statut_dome = 0).
duree_0	interval	Durée passée à conduire (statut_dome = 0).
nb_points_gps_1	bigint	Points GPS agrégés au 2 min (statut_dome = 1).
duree_1	interval	Durée passée à conduire (statut_dome = 1).
nb_points_gps_2	bigint	Points GPS agrégés au 2 min (statut_dome = 2).
duree_2	interval	Durée passée à conduire (statut_dome = 2).
nb_points_gps_7	bigint	Points GPS agrégés au 2 min (statut_dome = 7).
duree_7	interval	Durée passée à conduire (statut_dome = 7).
sridu	character varying(10)	Identifiant du principal secteur de recensement traversé.
smid	numeric	Identifiant du principal secteur municipal traversé.
arrid	character varying(10)	Identifiant du principal arrondissement traversé.
liste_sridu	character varying[]	Liste des identifiants des secteurs de recensement traversés.
zone_rechargement	integer	Identifiant de la principale zone de rechargement traversée.
couverture	integer	Identifiant de la période d'évolution du réseau Téo Taxi.

Figure 3-13 : Table GPS aux 2 minutes

3.7 Automatisation du traitement de données et optimisation

Les logiciels de traitement de données utilisés pour ce projet sont, dans l'ordre d'utilité, PostgreSQL version 1.22.1 (<https://www.postgresql.org>), Excel version Office 365 (<http://microsoft.com/excel>) et R version 1.0.136 (<https://www.r-project.org>). L'automatisation du traitement a été réalisée sous PostgreSQL et la majorité des graphiques sous Excel. Ce dernier permet une visualisation rapide et sophistiquée des données.

PostgreSQL est un logiciel libre dont la communauté d'utilisateurs est importante du fait des nombreuses informations et tutoriels disponibles sur internet. L'utilisation des requêtes SQL permet une grande liberté de traitement sur les tables de données. Cependant faire des requêtes SQL sur des données massives, comme cela peut l'être sur des tables GPS, peut engendrer des temps de calcul très longs de l'ordre de plusieurs jours si les requêtes codées ne sont pas optimisées. Ainsi l'un des enjeux d'un analyste de données est de gagner en temps de calcul afin d'obtenir le plus rapidement possible les résultats qu'il recherche. Voici les différentes manières d'augmenter les performances de PostgreSQL :

- Il est possible d'indexer des attributs gourmands en temps de calcul. Un index est une liste de données qui ont été ordonnées permettant dès lors un accès plus rapide à la donnée recherchée. Ainsi pour la table GPS, les attributs *temps*, *vehicle_id*, *driver_id* et *position_taxi* ont été indexés. La création d'index se fait à l'aide de la clause CREATE INDEX. Chaque index est associé à une table.
- Pour ajouter un attribut à une table, la recréer est plus rapide que de la mettre à jour. Cependant après la recréation d'une table, les index doivent être recréés aussi. Ainsi, s'agit-il de bien juger de s'il vaut mieux utiliser la clause CREATE TABLE ou UPDATE. La clause UPDATE permet de faire des opérations ligne par ligne, tandis que CREATE TABLE permet de faire des opérations attribut par attribut, c'est pourquoi cette dernière clause peut accélérer les temps de calcul.
- Plutôt que de modifier une table en plusieurs requêtes, il est préférable d'agréger l'ensemble de ses requêtes en une seule.
- PostgreSQL est beaucoup plus rapide lorsqu'il traite les enregistrements GPS d'une journée plutôt que les enregistrements d'un mois. Par exemple, pour le mois d'octobre, il vaut mieux appliquer la procédure de données 31 fois (une fois pour chaque jour) plutôt qu'une seule fois sur 1 mois.

Une fois les données traitées à l'aide du logiciel PostgreSQL, Excel est utilisé pour la visualisation. La simplicité d'usage et l'efficacité de la mise en forme visuelle permettent d'être rapide dans la création des différents graphiques.

3.8 Indicateurs

Les indicateurs permettent de décrire chaque objet composant le système de transport par taxi électrique durant une période de temps et un espace donnés. Ils permettent de donner un portrait de l'offre, en qualifiant par exemple la flotte de taxis et ses spécificités, l'équipe de chauffeurs et leurs caractéristiques, etc. Les indicateurs permettent aussi à l'opérateur de mieux connaître la demande et la réponse que donne l'offre en retour. Ainsi peut-il juger de la performance de la réponse apportée et l'ajuster selon son besoin. Les indicateurs permettent de donner un portrait de l'utilisation et de la performance du système de transport par taxi électrique, donnant alors à l'opérateur les moyens de prendre des décisions pour améliorer ou conserver l'offre de transport actuel.

3.8.1 Formes de visualisation utilisées

Quatre formes de visualisation ont été effectuées à l'aide de l'outil de visualisation des données : la statistique descriptive, la répartition temporelle, la cartographie et la distribution fréquentielle.

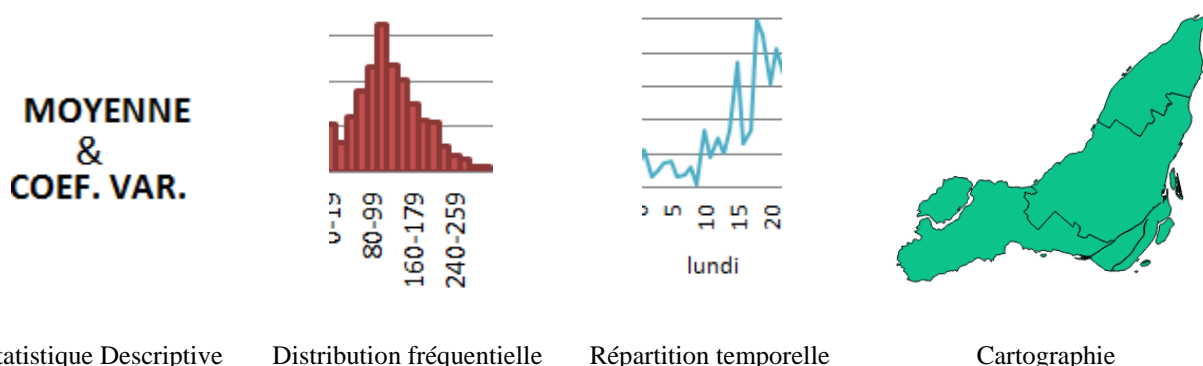


Figure 3-14 : Illustration des formes de visualisation utilisées

La statistique descriptive permet de décrire un objet en particulier en donnant une information générale à partir de plusieurs données particulières. C'est-à-dire que plutôt que de donner des informations sur un chauffeur en particulier, la statistique descriptive décrira un ensemble de chauffeurs. Des exemples de statistiques descriptives sont la moyenne, l'écart-type et le coefficient de variation.

La distribution fréquentielle se base sur le découpage d'un objet en plusieurs catégories. Par exemple, l'objet *course* peut être partagé en deux catégories : celles de *courses réalisées* et celles

de *courses annulées*. Ou encore l'objet *distance* peut être réparti en plusieurs intervalles de distance. L'histogramme ou les diagrammes circulaires peuvent être utilisés pour représenter une distribution fréquentielle.

La répartition temporelle permet de visualiser les différentes valeurs que prend un indicateur dans le temps. La courbe est la forme de représentation la plus usitée dans ce cas. Un exemple d'utilisation est l'évolution du nombre de courses par heure de la journée, pour une semaine moyenne d'un mois donné. La cartographie peut être complémentaire à la répartition temporelle, car quand cette dernière traduit l'évolution d'un objet dans le temps, la cartographie traduit son évolution dans l'espace.

La cartographie permet donc de visualiser les différentes valeurs que prend un indicateur dans l'espace sur un territoire donné. Un découpage par zone permet de comparer, par zone, les valeurs que prend un indicateur et ainsi de percevoir si un phénomène spatial se concentre sur une zone ou au contraire est uniforme sur le territoire. Le zonage le plus utilisé pour ce projet est le secteur de recensement de Statistiques Canada.

3.8.2 Présentation des indicateurs

Nous pouvons classer les indicateurs en deux catégories : les indicateurs d'offre et ceux de demande. Les indicateurs d'offre traitent des objets suivants : les chauffeurs et les véhicules, tandis que les indicateurs de la demande traitent des courses (réalisées et annulées) et des usagers.

De plus, parmi les indicateurs, certains servent à décrire la mise en service de système de transport par taxi et d'autres serviront à l'opérateur pour évaluer sa performance. Les indicateurs servant à évaluer la performance (en gras dans le tableau 3-3) vont permettre d'identifier si l'offre est optimisée ou non. Ce projet ne vise pas à « décider » quels sont les critères de performance les plus pertinents pour un réseau de taxis électriques. Cependant, des indicateurs d'aide à la décision ont été développés pour évaluer la performance du système de transport. Un système de transport par taxi est optimisé, ou performant, lorsque l'offre est en adéquation avec la demande et que les ressources investies pour l'établissement de l'offre sont minimales.

Les indicateurs de description quant à eux permettent de faire un suivi analytique du système de transport, ils vont décrire l'aspect quantitatif de l'utilisation de l'offre et donner un portrait global de l'utilisation du service de transport de Téo Taxi.

Cette présentation sera suivie d'une mise en application au chapitre 4 de ce présent mémoire. Il sera appliqué à la base de données de Téo Taxi les indicateurs fabriqués ci-dessous.

Tableau 3-3 : Tableau résumé des indicateurs fabriqués

OFFRE		
CHAUFFEURS	Statistiques descriptives	Revenu total par chauffeur par jour. Revenu par jour en fonction des types d'usagers et des types de services. Durée d'un quart de travail de jour. Distance parcourue par chauffeur par jour et par mois (avec et sans client). Nombre de chauffeurs par jour de semaine. Nombre de taxis par chauffeur.
	Répartition temporelle	Début des quarts de travail pour une semaine moyenne. Fin des quarts de travail pour une semaine moyenne.
	Distribution fréquentielle	Distribution des distances parcourue par jour par chauffeur + Moyenne de la distance sans client.
VÉHICULES	Statistiques descriptives	Nombre de taxis par jour de semaine. Distance parcourue par taxi par mois et par jour (avec et sans client). Durée de travail par taxi par mois et par jour (avec et sans client).
	Rép. temp.	Disponibilité et taux d'utilisation des taxis pour une semaine moyenne.

Tableau 3-3 : Tableau résumé des indicateurs fabriqués (suite)

	Cartographie	Densité horaire de disponibilité par secteur de recensement (h/km²) par période d'une journée moyenne (tous types de services confondus)
	Dist. fréq.	Distribution des distances parcourue par jour par taxi + Moyenne de la distance sans client.
BORNES	Cartographie	Nombre de zones de recharge. Localisation des zones de recharge.
DEMANDE		
COURSES REALISEES	Statistiques descriptives	Nombre de courses par chauffeur. Nombre de courses réalisées par jour de semaine. Distance moyenne d'une course. Durée moyenne d'une course. Distance d'une course par période de la journée. Durée d'une course par période de la journée.
	Rép. temp.	Évolution du nombre de courses par heure-chauffeur travaillée. Comparaison des longueurs et vitesses moyennes pour une semaine moyenne. Pourcentage de courses réalisées par types de services et de clients pour une semaine moyenne.

Tableau 3-3 : Tableau résumé des indicateurs fabriqués (suite)

	Distribution fréquentielle	<p>Distribution des chauffeurs selon le nombre de courses effectuées par jour.</p> <p>Distribution cumulée du nombre de courses, des revenus générés, des distances et des durées de courses en fonction du pourcentage cumulé des chauffeurs.</p> <p>Distribution des distances des courses par période de la journée.</p> <p>Distribution des durées des courses par période de la journée.</p>
	Cartographie	Densité horaire d'origines et de destinations par secteur de recensement par période d'une journée moyenne (tous types de services confondus).
COURSES ANNULÉES	Stat	Nombre de courses annulées par jour.
	Répartition temporelle	<p>Comparaison du nombre de courses réalisées et du nombre de courses annulées pour une semaine moyenne.</p> <p>Raisons des annulations pour une semaine moyenne.</p> <p>Pourcentage de courses annulées par types de services et de clients pour une semaine moyenne.</p>
	Carto.	Densité horaire d'annulations par secteur de recensement par période d'une journée moyenne (tous types de services confondus).
USAGERS	Statistique Descriptive.	<p>Temps d'attente pour une course réalisée.</p> <p>Nombre d'utilisateurs par jour.</p> <p>Nombre de courses réalisées par jour, par utilisateur actif.</p> <p>Nombre de courses annulées par jour, par utilisateur actif.</p>

Tableau 3-3 : Tableau résumé des indicateurs fabriqués (suite)

	Répartition temporelle	Temps d'attente pour une semaine moyenne. Pourcentage d'utilisateurs actifs ayant annulé au moins une fois leurs courses pour une semaine moyenne
	Distribution fréquentielle	Distribution du nombre d'annulations par nombre d'utilisateurs. Distribution cumulée du nombre de courses en fonction du pourcentage cumulé d'utilisateurs. Longueur moyenne d'une course réalisée en fonction du temps d'attente de l'utilisateur. Nombre cumulé de courses annulées en fonction du temps d'annulation.

La fabrication des indicateurs est présentée ci-dessous. Il s'agit de donner un portrait des objets étudiés et de rendre compte de l'histoire vécue par chacun d'eux à travers différentes descriptions et évaluations. Leur fabrication se base sur les données partagées par la compagnie de taxis électriques Téo Taxi.

L'objet *poste d'attente* n'est pas approfondi dans les analyses, car les utilisateurs de Téo Taxi n'utilisent pas les postes d'attente. Ceux-ci commandent leurs courses à des points d'origine où les chauffeurs les rejoignent. Lacombe (2016) montre que, lorsque les postes d'attente sont utilisés par les utilisateurs, les plus utilisés ne sont pas nécessairement ceux situés au centre-ville, mais plutôt ceux stratégiquement localisés sur le territoire. Elle montre que les postes d'attente localisés à proximité des stations de métro sont les plus utilisés, démontrant par conséquent la complémentarité entre le métro et le taxi.

Cette étude sur les indicateurs de description et d'aide à l'évaluation de la performance d'un système de transport par taxi électrique reprend et fait suite à l'étude menée par Lacombe (2016).

3.8.3 Indicateurs d'offre

3.8.3.1 Indicateurs des chauffeurs

La journée des chauffeurs est décrite par leurs quarts de travail et leurs activités durant cette journée. Le nombre de chauffeurs, leur revenu moyen, la distance qu'ils parcourent en moyenne avec ou sans client sont autant d'indices sur leur activité et la performance de l'offre.

Indicateurs de quarts de travail

Les débuts et fins des quarts de travail sont calculés pour une semaine moyenne d'un mois donné. Pour chaque jour de semaine et chaque heure de la journée, le nombre de chauffeurs commençant et terminant son quart de travail est compté.

La méthode de calcul employée ne tient pas compte du nombre de débuts de quarts de travail entre minuit et 1h du matin, et du nombre de fins de quarts entre 23h et minuit. Il existe des chauffeurs qui commencent à travailler un jour j et finissent leur quart le jour $j+1$. Compter, avec cette méthode, le nombre de débuts de quart de travail entre minuit et 1h du matin reviendrait à compter des quarts de travail qui peuvent avoir débuté la veille. De même, entre 23h et minuit, cela reviendrait à compter des quarts de travail qui peuvent se finir le lendemain. Cette méthode peut donc être peaufinée en bordure de journée.

La durée d'un quart de travail de jour donne la durée moyenne d'un quart de travail pour un chauffeur ayant commencé et fini son quart durant la journée même. C'est pourquoi ce sont des quarts de travail *de jour*. Les chauffeurs ayant travaillé à cheval sur deux jours ne sont pas comptés.

Indicateurs de distance parcourue

La distance parcourue par chauffeur par jour et par mois permet de mesurer la différence de distance passée à conduire avec un client et celle passée à conduire sans client. Pour plus de précision, nous sommes amenés à fabriquer la distribution fréquentielle suivante :

La distribution des distances parcourue par jour par chauffeur auquel il est ajouté, pour chaque fréquence, *la moyenne de la distance sans client*. Cet indicateur de performance permet d'évaluer, pour chaque groupe de distance parcourue dans la journée, la distance moyenne passée à conduire sans client.

Autres indicateurs

Le revenu total par chauffeur par jour est calculé en ne prenant que les valeurs du taximètre > 1 . Cette valeur arbitraire permet de ne pas prendre en compte les courses où le revenu est nul. Puis, le revenu peut aussi se calculer en fonction du type d'usager (corporatif ou particulier) ou type de service (aéroportuaire ou régulier).

Le nombre de chauffeurs par jour et *le nombre de taxis par chauffeurs* sont deux autres indicateurs de description.

3.8.3.2 Indicateurs des véhicules

Le nombre de véhicules, la distance qu'ils parcourent, le temps durant lequel ils sont utilisés permettent de décrire l'activité de la flotte. Des indicateurs de performance sont calculables à l'aide de leur statut de disponibilité.

Indicateurs de distance et de durée

Comme pour les chauffeurs, il est possible de calculer *la distance parcourue par taxi par jour et mois, avec ou sans client*. Ces statistiques descriptives s'accompagnent aussi d'une *distribution fréquentielle des distances parcourues par jour par taxi*.

Comme un quart de travail est la durée de travail d'un chauffeur, il est possible de calculer le temps de travail d'un véhicule à l'aide des indicateurs suivants : *les durées de travail par taxi par mois et par jour, avec et sans client*.

Indicateurs de disponibilités temporelle et spatiale

Il est possible de calculer la *disponibilité* et le *taux d'utilisation des taxis pour une semaine moyenne* d'activité à l'aide de la table GPS. En effet un des attributs de la table est le *statut_dome* indiquant à chaque enregistrement si le véhicule est disponible (*statut_dome* = 2), indisponible, mais pas en course (*statut_dome* = 1), en course (*statut_dome* = 0) et en train de se connecter au dispositif GPS (*statut_dome* = 7). Ainsi à chaque minute de la journée, nous sommes capables de suivre le statut de disponibilité de chaque taxi. En particulier, il est possible de compter *le nombre de véhicules disponibles* à chaque 10 minutes de la journée.

Un véhicule disponible est un véhicule dans l'attente d'un client et qui, donc, soit conduit à vide, soit est à l'arrêt. Le but d'un opérateur est de minimiser ce nombre tout le long de la journée. Le *taux d'utilisation des taxis* est le nombre total de véhicules avec client sur le nombre total de véhicules disponibles. Ce taux peut être calculé pour une semaine moyenne d'un mois donné. Il donne un portrait des heures de la journée où la flotte de taxi est le plus, ou le moins, utilisée.

L'évolution de la disponibilité de la flotte des taxis par heure de la journée peut être complétée par une analyse de son évolution spatiale. Un indicateur spatial de disponibilité permet de connaître les lieux où, durant la journée, les taxis sont les plus disponibles ou au contraire le moins. Cet indicateur se présente sous forme de cartes. Il donne la *densité horaire de disponibilité par secteur de recensement (h/km^2) par période d'une journée moyenne* (tous types de services confondus). Pour le fabriquer, il s'agit de construire une carte de disponibilité par période de la journée. Le découpage périodique choisi est le suivant : 0-6h, 6-9h, 9-15h, 15-18h, 18-24h. Ce découpage permet de distinguer les périodes de pointes (6-9h et 15-18h) des périodes habituellement creuses (0-6h, 9-15h, 18-24h). Pour chaque période de la journée et pour chaque secteur de recensement, on compte le nombre d'heures de disponibilité de la flotte de véhicules. On divise ce nombre par l'aire du secteur et le nombre d'heures de la période horaire.

3.8.3.3 Indicateurs des sites de bornes

Les données portant sur l'activité des bornes de recharge nous permettent à ce jour de localiser les sites sur lesquels elles sont placées. Une cartographie des zones de recharge est donc un indicateur permettant de visualiser à la fois *le nombre de sites de bornes et leur localisation sur le territoire*.

3.8.4 Indicateurs de demande

3.8.4.1 Indicateurs des courses réalisées

Indicateurs liés aux chauffeurs

Il est possible de calculer, pour une période donnée, le *nombre de courses par chauffeur*. La *distribution des chauffeurs selon le nombre de courses effectuées par jour* donne le nombre de courses réalisées en fonction du pourcentage de chauffeurs-jour. Cet indicateur donne, pour une

journée moyenne donnée, le pourcentage du total de chauffeurs ayant effectué 1 course, 2 courses, 3 courses, etc.

Un indicateur de performance est donné par la *distribution cumulée du nombre de courses en fonction du pourcentage cumulé des chauffeurs*. Cette distribution est obtenue de la manière suivante : il faut classer les chauffeurs selon l'ordre décroissant du nombre de courses réalisées (cancellation_reason_code = 0) durant une période donnée. Par exemple, pour le mois d'octobre 2016, 50% des courses ont été réalisées par 31% des chauffeurs. Et 50% des chauffeurs ont fait près de 72% des courses.

De manière similaire, il est possible de fabriquer les distributions suivantes : *la distribution cumulée du revenu généré en fonction du pourcentage cumulé des chauffeurs*, *la distribution cumulée de la distance et de la durée des courses en fonction du pourcentage cumulé des chauffeurs*.

Un autre indicateur de performance est donné par *l'évolution du nombre de courses par heure-chauffeur travaillée*. Pour obtenir ce nombre pour chaque jour d'un mois, il compte le nombre de courses réalisées chaque jour et de le diviser par le nombre total d'heures travaillées par l'ensemble des chauffeurs cette journée.

Indicateurs de débuts et fin de courses

Cette analyse temporelle est corroborée par une analyse spatiale des lieux d'origine et de destination des courses. L'indicateur associé est le suivant : *la densité horaire d'origines et de destinations par secteur de recensement par période d'une journée moyenne* (tous types de services confondus). Il se fabrique en comptant le nombre d'extrémités de courses par secteur de recensement. Ce nombre doit ensuite être divisé par l'aire du secteur et le nombre d'heures de la période horaire.

Indicateurs de distance et de durée

Une course se caractérise par sa longueur et sa durée. Ainsi différents indicateurs peuvent être fabriqués : *la distance d'une course*, *la durée d'une course*. Puis il est possible d'observer la variation de cette *distance et de cette durée par période de la journée* : 0-6h, 6-9h, 9-15h, 15-18h, 18-24h.

Puis pour une analyse plus fine encore, la *distribution des distances des courses par période de la journée* et la *distribution des durées des courses par période de la journée* sont calculables. Il s'agit de construire plusieurs groupes de distances et de durées et de mettre dans chaque groupe le pourcentage de courses de la journée adéquat.

Nous pouvons comparer pour une semaine moyenne, les longueurs et les vitesses moyennes de courses. Ainsi pour chaque journée de la semaine, il est possible d'observer la variation horaire de la longueur moyenne d'une course et de la comparer à la variation horaire de vitesse.

Type de courses réalisées

Le *pourcentage de courses réalisées par type de service et de client pour une semaine moyenne* donne le pourcentage de courses réalisées pour un service aéroportuaire ou régulier (`service_type = 'airport'` ou `service_type = 'regular'`) quand l'utilisateur est un particulier ou un représentant de corporation (`client_type = 0` ou `client_type = 1`).

Autres indicateurs

Un autre indicateur à ajouter aux précédents est celui-ci : le *nombre de courses réalisées par jour de semaine* qui est un indicateur de description.

3.8.4.2 **Indicateurs des courses annulées**

Raisons des annulations

Il existe 5 raisons possibles pour qu'une course soit annulée : l'absence de client, le temps d'attente est trop long (annulation client), le client a changé d'avis (annulation client), la raison est inconnue (annulation client) ou autre raison (annulation Téo) respectivement données par le `cancellation_reason_code = 1, 2, 3, 4 et 5`. Le *pourcentage de raisons des annulations pour une semaine moyenne* permet de mettre en avant la raison d'annulation principale chaque heure de la journée, pour chaque jour de la semaine.

3.8.4.2.1 *Types de courses annulées*

Comme pour les courses réalisées, il est possible de construire un indicateur caractérisant le pourcentage des courses annulées en fonction de leur type de service, ou de client servi.

L'indicateur de description fabriqué donne le *pourcentage de courses annulées par types de services et de clients pour une semaine moyenne*.

Autres indicateurs

Le *nombre de courses annulées par jour de semaine* est un autre indicateur de description de base. L'évolution du nombre de courses annulées chaque heure de la journée, pour une semaine moyenne d'un mois donné, peut être confrontée à l'évolution du nombre de courses réalisées. Le *taux de courses réalisées* (= nombre de courses réalisées / (nombre de courses réalisées + nombre de courses annulées)) est un indicateur de performance.

Encore une fois, l'évolution temporelle du nombre de courses annulées se corrobore avec une analyse de son évolution spatiale. La *densité horaire d'annulations par secteur de recensement par période d'une journée moyenne* (tous types de services confondus) permet de visualiser l'évolution du phénomène sur le territoire. Pour fabriquer cet indicateur, il faut calculer le nombre de courses annulées par secteur de recensement. On divise ce nombre par l'aire du secteur et le nombre d'heures de la période horaire.

3.8.4.3 Indicateurs des usagers

Les usagers du taxi forment la demande, nous pouvons décrire leur comportement individuel ou de groupe à l'aide de différents indicateurs présentés ci-dessous. Nous pouvons distinguer l'*usager actif* de l'*usager*. Tous les usagers de Téo Taxi ne font pas de courses tous les jours. Ainsi il a été nécessaire de poser le concept d'*usager actif*. Un usager actif est un usager ayant fait une course durant une journée.

Indicateur lié aux courses réalisées

Le temps d'attente d'un usager est défini comme étant la différence entre l'instant de réservation de la course et l'instant de départ au lieu d'origine. ($\text{client_waiting_time} = \text{origin_departure_time} - \text{booking_time}$). Le calcul du *temps d'attente moyen d'un usager pour une semaine moyenne* est un indicateur de performance. Pour chaque jour de la semaine, et à chaque heure de la journée, le temps d'attente moyen pour une course est donné.

La *longueur moyenne d'une course réalisée en fonction du temps d'attente de l'usager* permet de mettre en avant le temps d'attente moyen qu'un usager est prêt à passer pour attendre sa course

dépendamment de la longueur de cette dernière. Un temps d'attente moyen est calculé pour différentes longueurs de courses (0-1 km, 1-2 km, etc.).

Afin de connaître les usagers les plus fidèles, il est possible de construire l'indicateur suivant : *la distribution cumulée du nombre de courses en fonction du pourcentage cumulé des usagers*. Ainsi durant la période étudiée, nous pouvons connaître quel pourcentage du total des usagers fait quel pourcentage du total des courses. Cet indicateur se construit comme *la distribution du nombre de courses en fonction du pourcentage cumulé des chauffeurs*.

Indicateurs liés aux courses annulées

Le temps d'annulation est défini comme étant la différence entre l'instant où l'utilisateur annule sa course et l'instant où il l'avait commandé ($\text{delay_before_cancellation_seconds} = \text{cancellation_time} - \text{booking_time}$). Le nombre cumulé de courses annulées en fonction du temps d'annulation indique combien de courses sont annulées au bout de combien de temps. Pour une distribution sur le temps d'annulation (0-1 min, 1-2 min, etc.), le nombre de courses annulées ($\text{cancellation_reason_code} = 1, 2, 3 \text{ ou } 4$) est compté pour chaque intervalle. La raison d'annulation n°5 n'est pas prise en compte puisque l'utilisateur n'en est pas à l'origine, l'annulation vient de Téo Taxi.

La *distribution du nombre d'annulations par nombre d'utilisateurs* donne, pour une période donnée, le nombre d'utilisateurs ayant annulé 1 course, le nombre de ceux ayant annulé 2 courses, 3 courses, etc. ($\text{cancellation_reason_code} = 1, 2, 3 \text{ ou } 4$).

En calculant le *pourcentage d'utilisateurs ayant annulé au moins une fois leurs courses pour une semaine moyenne*, nous pouvons connaître le pourcentage du total des utilisateurs ayant annulé au moins une course chaque jour. ($\text{cancellation_reason_code} = 1, 2, 3 \text{ ou } 4$).

Autres indicateurs

Les statistiques descriptives suivantes ont aussi été calculées : *le temps d'attente pour une course réalisée, le nombre d'utilisateurs actifs par jour, le nombre de courses réalisées par jour par utilisateur actif, ainsi que le nombre de courses annulées par jour par utilisateur actif*.

CHAPITRE 4 APPLICATION : UTILISATION DES INDICATEURS

Le chapitre précédent a présenté les indicateurs construits pour décrire et évaluer l'offre et la demande en transport de taxis électriques. Il s'agit maintenant d'appliquer les indicateurs fabriqués aux données de Téo Taxi.

4.1 Cadre d'étude

Le cadre d'étude spatiotemporel est le suivant : le territoire d'analyse comprend l'île de Montréal. Durant l'année 2016, la couverture du réseau a demeuré au sein de la zone A-11 de taxis (cf. figure 3-1). La période temporelle d'étude commence le 1er juillet 2016 et finit le 31 décembre 2016. Les analyses distinguent 2 périodes correspondant aux saisons : l'été, allant du 1er juillet au 30 septembre, et l'automne, allant du 1er octobre au 31 décembre. En effet, tant les jours de vacances que les phénomènes climatiques ont un impact sur les réseaux de transport par taxis (Pelé et al., 2014). Cette distinction est d'autant plus pratique qu'elle permet de prendre en compte l'extension de la couverture du réseau ayant eu lieu le 14 septembre (cf. figure 3-3).

Tableau 4-1 : Résultats du prétraitement des données - table de courses & table GPS

Table courses		
Nombre de données :	Avant nettoyage : 262 484	0,9% des enregistrements supprimés
	Après nettoyage : 260 250	
Table GPS		
Nombre de données :	Avant nettoyage : 231 066 869	0,3% des enregistrements supprimés
	Après nettoyage : 230 395 852	
Table GPS aux 2 min		
Nombre de données :	Avant prétraitement : 230 395 852	96% des enregistrements agrégés
	Après prétraitement : 8 580 002	

Le traitement de données est appliqué à la table de courses et la table GPS fournies par Téo Taxi. Ainsi les tables ont été prétraitées et standardisées comme présenté dans les parties 3.3 et 3.4. Le Tableau 4-1 donne les nombres d'enregistrements contenus dans les tables courses et GPS après prétraitement des données.

Durant cette analyse, il faut avoir à l'esprit que Téo Taxi a été en perpétuelle évolution durant l'année 2016. Comme le montre la section 3.1, le nombre de chauffeurs et le nombre de véhicules ont continuellement augmenté entre juillet et décembre. La taille du réseau a connu une expansion. Même si nous ne pouvons pas savoir si c'est l'offre en taxi qui influence la demande ou inversement, nous constatons que l'augmentation de l'offre s'accompagne d'une augmentation de la demande. Par ailleurs, il est à noter que les tendances sont parfois suffisamment similaires entre l'été et l'automne pour ne plus distinguer ces deux périodes.

4.2 Application : Indicateurs d'Offre

Tableau 4-2 : Tableau d'indicateurs d'offre

	INDICATEUR		MOYENNE	COEF. VAR.
CHAUFFEURS	Revenu par chauffeur par jour		144,22\$	17%
	Revenu par chauffeur par jour	Usagers réguliers	132,54\$	17%
		Usagers corporatifs	32,73\$	17%
		Courses régulières	119,62\$	18%
		Courses aéroportuaires	119,86\$	20%
	Durée d'un quart de travail de jour		7,8h	50%
	Distance parcourue par chauffeur:			
	Par mois	Distance totale	1153 km	72%
		Distance sans client	657 km	73%
	Par jour	Distance totale	122 km	107%
		Distance sans client	69 km	117%
	Nombre de chauffeurs par jour			
	Lundi		90 chauffeurs	22%
	Mardi		104 chauffeurs	30%
	Mercredi		113 chauffeurs	30%
	Jeudi		125 chauffeurs	27%
	Vendredi		121 chauffeurs	30%
	Samedi		94 chauffeurs	42%
	Dimanche		81 chauffeurs	38%
	Nombre de taxis par chauffeur		1,6 veh.	47%
VÉHICULES	Nombre de taxis par jour			
	Lundi		81 veh.	24%
	Mardi		87 veh.	24%
	Mercredi		88 veh.	24%
	Jeudi		88 veh.	23%
	Vendredi		89 veh.	24%
	Samedi		87 veh.	25%
	Dimanche		82 veh.	26%
	Distance parcourue par taxi			
	Par mois	Distance totale	4078 km	54%
		Distance sans client	2375 km	50%
	Par jour	Distance totale	190 km	100%
		Distance sans client	111 km	105%
	Durée de travail par taxi			
	Par mois	Durée totale	359 h	41%
		Durée sans client	296 h	41%
	Par jour	Durée totale	16,7 h	10%
		Durée sans client	13,7 h	10%

Un chauffeur génère en moyenne un revenu de 144,20\$ par jour de travail. Le revenu attribuable aux usagers réguliers est 132,20\$. La part attribuable aux usagers corporatifs est 32,70\$. Les courses régulières lui permettent de gagner en moyenne près de 119,60\$ par jour. Et si sa journée est composée uniquement de courses aéroportuaires, le revenu est près de 120 \$.

Le nombre de chauffeurs par jour est de 90 pour un lundi moyen, monte à 125 chauffeurs en milieu de semaine et redescend petit à petit pour descendre assez brusquement à 81 chauffeurs le dimanche. La figure 3-3, section 3.1, rappelle que si l'équipe de chauffeurs Téo était à plus de 150 chauffeurs début juillet, ce nombre est passé à plus de 250 chauffeurs fin décembre. Un chauffeur utilise en moyenne 1,6 véhicules par jour, le coefficient de variation de 47% montre la grande variabilité du nombre de véhicules par chauffeur. En effet, l'interchangeabilité (*swapping*) des taxis fait qu'un chauffeur peut conduire jusqu'à plus de 2 véhicules différents durant une journée. La médiane du nombre de véhicules par chauffeur est à 1 et le 75e centile est à 2.

Un chauffeur, de jour, travaille en moyenne 7,8 h avec un coefficient de variation de 42%. Cette variation peut s'expliquer par le fait que les chauffeurs de taxi prennent des temps de pauses différents. Les figure 4-1 et figure 4-2 montrent respectivement les débuts et les fins de quarts de travail. C'est vers 6-7 h du matin que les débuts de quart de travail connaissent un pic qui peut monter à plus de 10 débuts de quarts l'été et plus de 15 l'automne. Un second pic, plus petit, existe en début d'après-midi entre 13 et 15 h selon les jours marquant probablement la fin des temps de pauses du repas de midi. Ce pic monte à 6-8 débuts de quarts l'été et 6-10 l'automne. Ces tendances sont invariantes, quel que soit le jour, sauf le dimanche où les débuts sont hétérogènes toutes la journée.

Quand aux fins de quarts de travail pour une journée ouvrable, 3 pics sont observables qui montent jusqu'à un peu plus de 6 fins l'été et 10 fins l'automne. L'un se trouve vers 14-15 h, le second entre 18-20 h qui se confond avec le dernier qui apparaît vers 22h. Il est remarquable que durant le samedi, l'heure du dernier pic se déplace. Il se déplace à la journée d'après, soit le dimanche, à 2h du matin. Ceci montre que les chauffeurs sont actifs les samedis soirs. En effet, nous constaterons plus tard que la demande de courses entre le samedi soir et le dimanche matin est élevée. Les tendances durant l'été et l'automne se ressemblent pour une journée ouvrable, mais pour la fin de semaine les tendances s'amplifient en automne par rapport à l'été. En effet, les pics

du matin et du soir doublent le samedi et le dimanche. L'effet de l'augmentation du nombre de chauffeurs engagés par Téo explique ce phénomène.

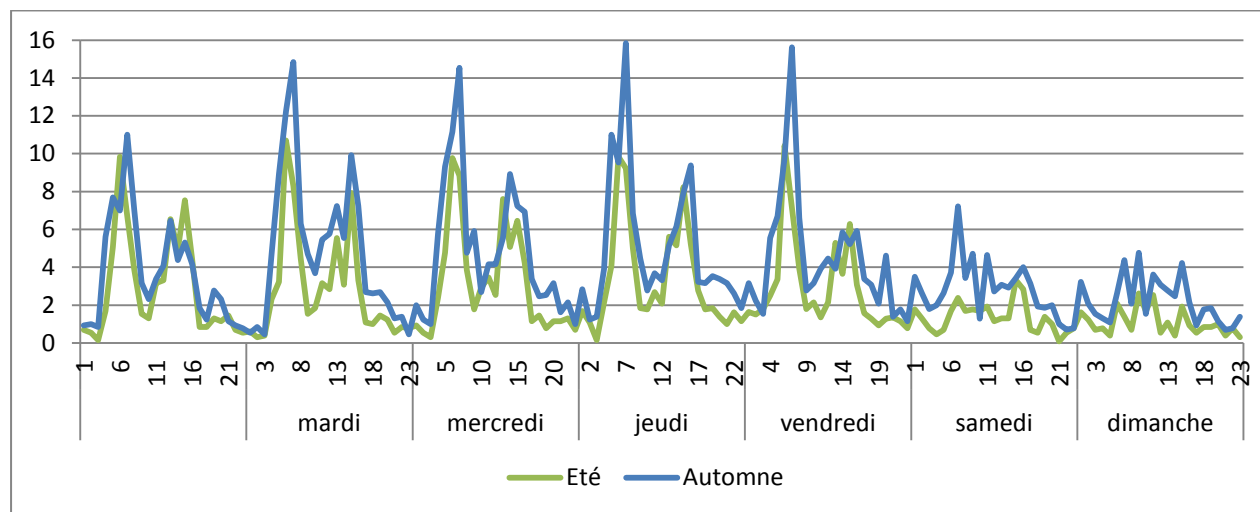


Figure 4-1 : Début des quarts de travail pour une semaine moyenne

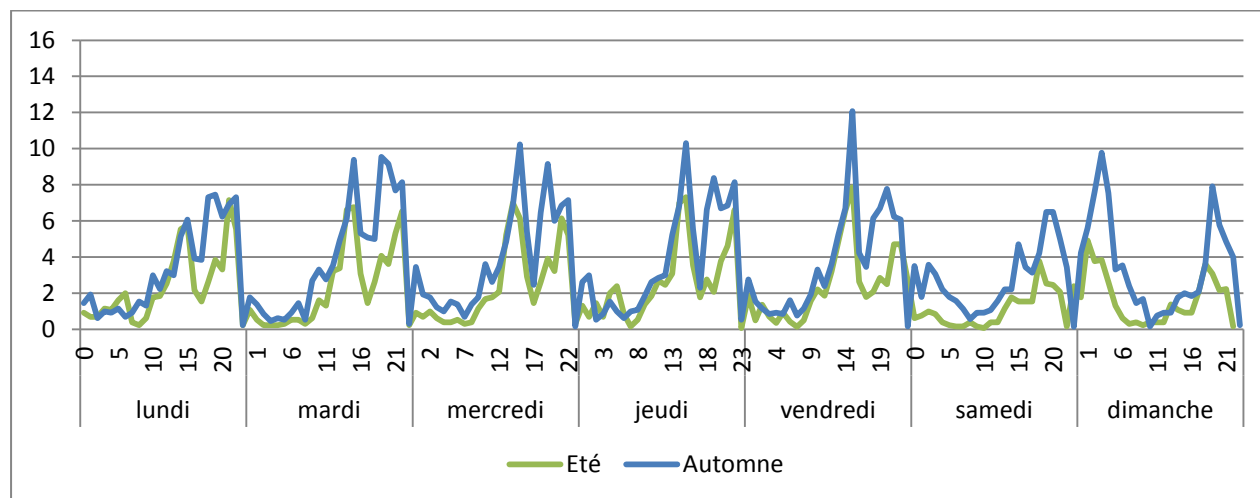


Figure 4-2 : Fin des quarts de travail pour une semaine moyenne

Par ailleurs, la distance parcourue chaque jour par un chauffeur est de 122 km au total et 69 km à vide. Pour été et automne confondus, la figure 4-3 montre que moins le chauffeur parcourt de distance, plus il roulera à vide. Et inversement : plus le chauffeur parcourt de distance, moins il roulera à vide. Pour les distances supérieures à 60 km, le pourcentage de la distance faite sans client est inférieur à 60%. Lorsque nous comparons la figure 4-3 et figure 4-4, nous trouvons des tendances similaires pour les véhicules. Comme les chauffeurs, les véhicules ayant parcouru des distances supérieures à 60 km ont un pourcentage de distance à vide inférieur à 65%.

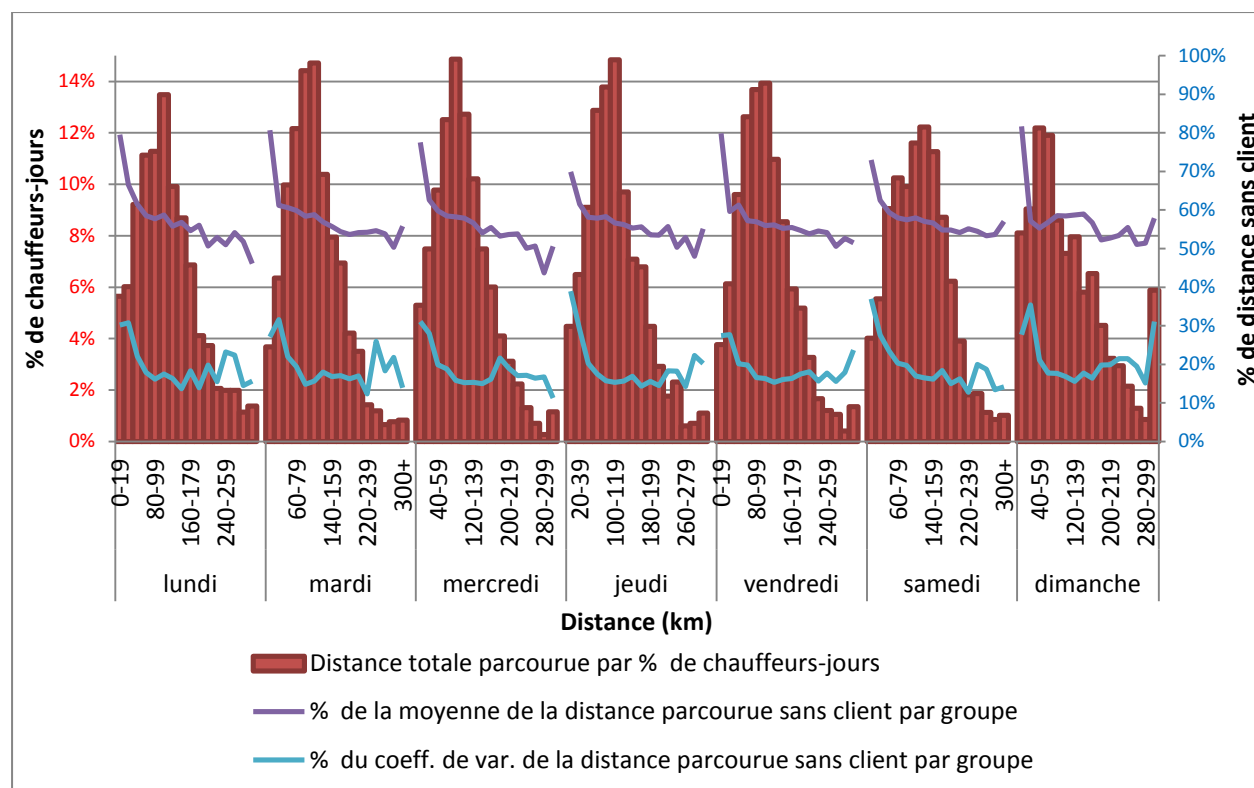


Figure 4-3 : Distribution des distances parcourues par jour par chauffeur et moyenne de la distance sans client - (Été + Automne)

Chaque jour, un taxi parcourt en moyenne des distances plus élevées qu'un chauffeur. Il parcourt en moyenne 190 km par jour au total et 111 km à vide.

Les distances parcourues par chauffeur ou par véhicule ont augmenté avec l'expansion de la couverture du réseau. Si nous ne considérons que les mois d'octobre à décembre, le nombre de kilomètres totaux parcourus par chauffeur passe à 1678 km par mois (dont 957 km à vide), et à 131 km par jour (74 km à vide). Elle augmente donc de 31% par mois et de 7% par jour. En ce qui concerne les véhicules, elle est de 5 331 km par mois (3 079 km à vide) et 205 km (118 km à vide). La distance augmente donc de 23% par mois et de 7% par jour si on ne considère que la période d'automne.

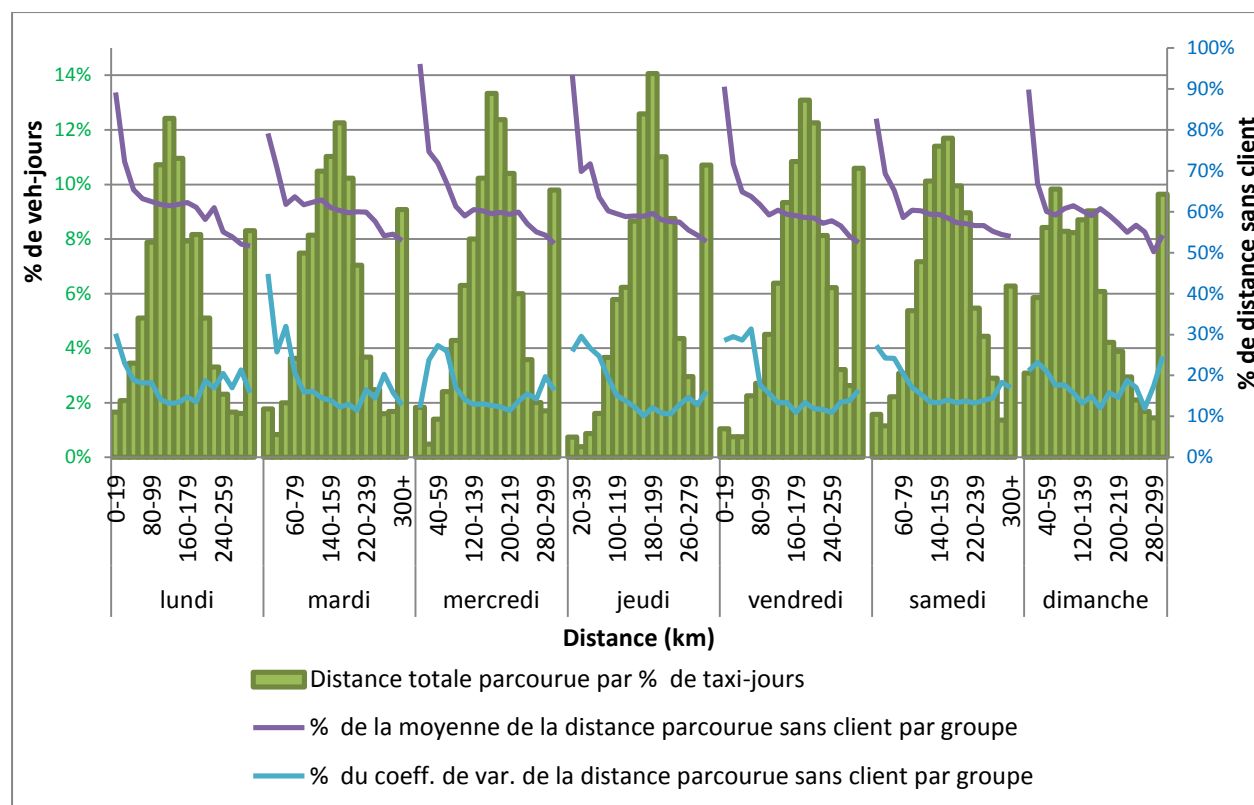


Figure 4-4 : Distribution des distances parcourues par jour par taxi et moyenne de la distance sans client - (Été + Automne)

Au cours de la semaine, le nombre moyen de véhicules actifs par jour tourne autour de 85. Bien que le nombre de véhicules dans la flotte ait augmenté au cours des mois, la variabilité est faible (autour de 25% chaque jour). Ce nombre est moindre le lundi (81 véhicules) et le dimanche (82 véhicules) et atteint son maximum le vendredi avec 89 véhicules. L'offre (chauffeurs et véhicules) augmente chaque jour jusqu'à jeudi-vendredi car la demande est plus élevée ces jours-là, comme nous le verrons dans l'analyse des courses. Ceci nous amène à nous intéresser au taux d'utilisation de la flotte de taxis électriques avec la figure 4-5. Le taux d'utilisation de la flotte est le rapport du nombre de taxis en course (statut_dome = 0) sur le nombre de taxis disponibles (statut_dome = 1).

Les profils de disponibilité d'été et d'automne se ressemblent. D'après la figure 4-5, chaque jour de semaine le taux d'utilisation varie globalement entre 30 et 70% selon le jour et l'heure de la journée. Et chaque jour, en milieu de journée ce nombre dépasse globalement les 50%, c'est-à-dire que plus d'un véhicule sur deux est en course. Le taux d'utilisation connaît des pics durant les heures de pointe (6-9 h ou 15-18 h), mais aussi durant le soir. Ceci se voit clairement pour les

journées de jeudi et vendredi : ces journées-là, il existe un pic commençant vers 22 h et pouvant se terminer à minuit ou plus tard. De plus, le taux d'utilisation est le plus faible (inférieur à 40%) le samedi et le dimanche très tôt le matin entre 4 h et 8 h du matin. Lorsque nous analyserons la demande, nous verrons que bien que le taux d'utilisation est faible (inférieur à 40%), le taux de courses réalisées est élevé (supérieur à 70%) durant ces heures de fin de semaine.

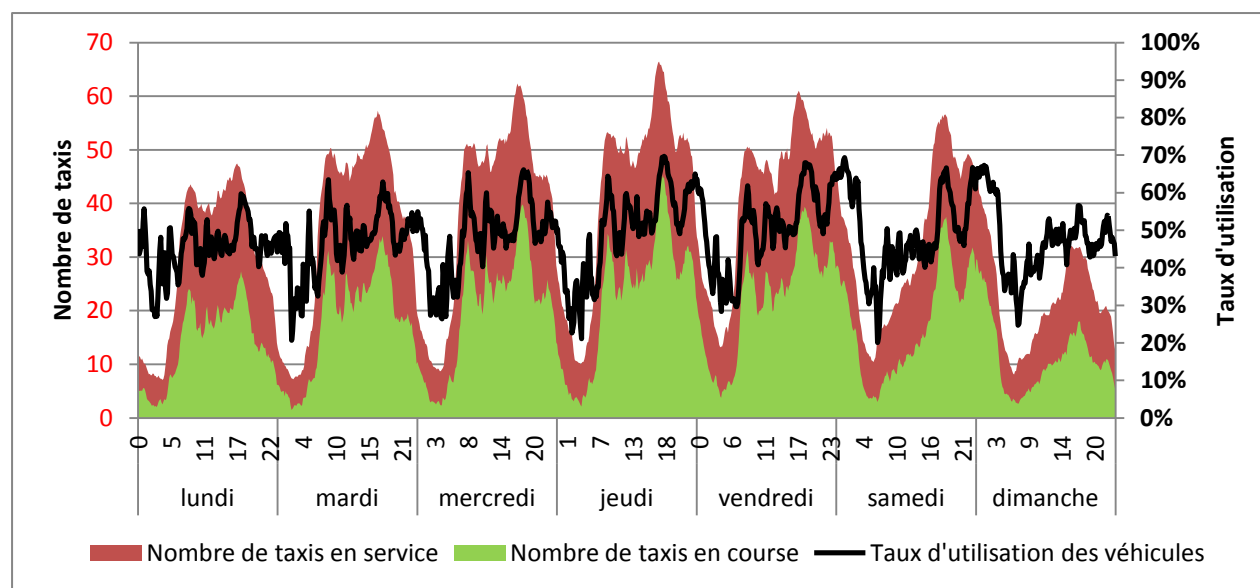
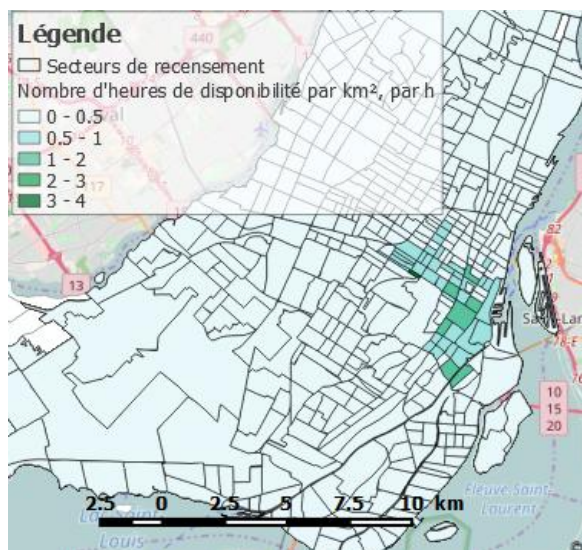


Figure 4-5 : Disponibilité et taux d'utilisation des taxis pour une semaine moyenne - (Été + Automne)

Pour compléter l'analyse précédente, les figures suivantes montrent l'évolution spatiale de la densité de disponibilité de taxis par période pour une journée moyenne d'été (figure 4-6) et d'automne (figure 4-7). Pour une période et un secteur de recensement donnés, une densité de 1 h/km² équivaut à 1 h de taxi disponible par km² pour ce secteur et pour cette période. La densité de disponibilité est la plus dense en centre-ville et sur le Plateau Mont-Royal, de 15 à 18h pour l'été et l'automne. En automne, il y a augmentation du nombre d'heures de taxis disponibles tout au long de la journée. L'augmentation du nombre de véhicules dans la flotte de taxis peut expliquer ce phénomène. Cette densité monte à 3-4 h/km² de disponibilité proche de St-Patrick, mais aussi en centre-ville en automne plus particulier. Les taxis disponibles semblent donc se regrouper au garage de Téo Taxi sur St-Patrick et sur la zone de recharge de la Caisse de Dépôt en centre-ville. Par ailleurs il est notable, malgré l'expansion du réseau, que les taxis disponibles restent concentrés autour du centre-ville.



(a) - 0-6h



(b) - 6-9h



(c) - 9-15h

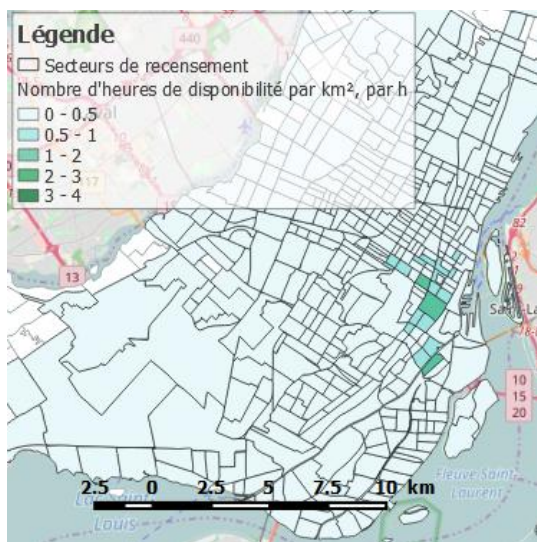


(d) - 15-18h

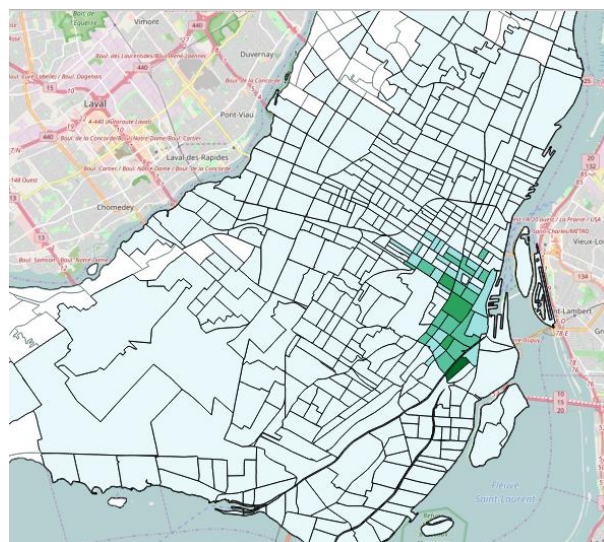


(e) - 18-24h

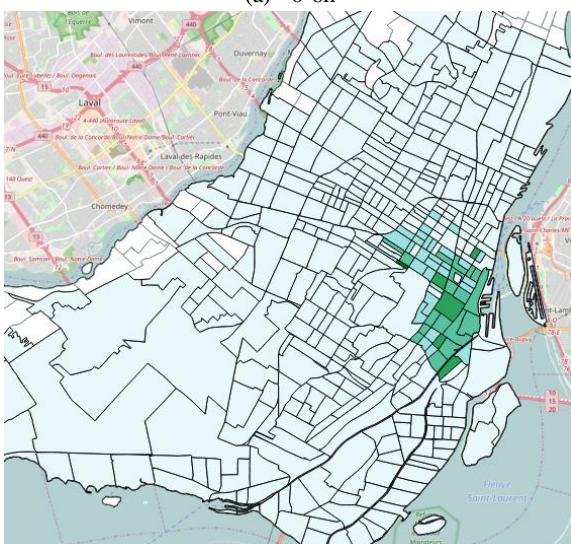
Figure 4-6 : Densité horaire de disponibilité par secteur de recensement (h/km²) pour une journée moyenne (tous types de services confondus) entre (a) 0-6h, (b) 6-9h, (c) 9-15h, (d) 15-18h, (e) 18-24h. - Été



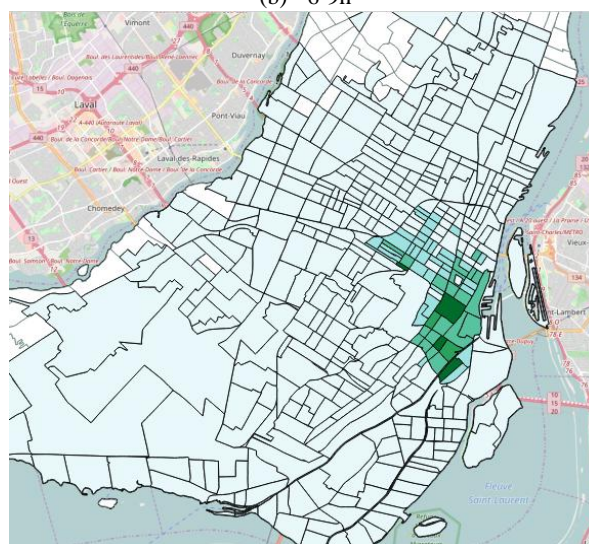
(a) - 0-6h



(b) - 6-9h



(c) - 9-15h



(d) - 15-18h



(e) - 18-24h

Figure 4-7 : Densité horaire de disponibilité par secteur de recensement (h/km²) pour une journée moyenne (tous types de services confondus) entre (a) 0-6h, (b) 6-9h, (c) 9-15h, (d) 15-18h, (e) 18-24h. - Automne

4.3 Application : Indicateurs de Demande

Tableau 4-3 : Tableau d'indicateurs de demande

	INDICATEUR					MOYENNE	COEF. VAR.
COURSES REALISEES	Nombre de courses par chauffeur					8,9 courses	18%
	Nombre de courses réalisées par jour						
	Lundi					788 courses	36%
	Mardi					995 courses	38%
	Mercredi					1115 courses	35%
	Jeudi					1259 courses	33%
	Vendredi					1250 courses	37%
	Samedi					1019 courses	41%
	Dimanche					645 courses	36%
	Distance d'une course					6,3 km	173%
	Durée d'une course					16,5 min	148%
	Indicateur		0-6h	6-9h	9-15h	15-18h	18-24
	Dist	moy	8,8 km	6,1 km	5,8 km	6,5 km	5,8 km
		C.V	310%	92%	100%	94 %	89%
		med	5,4 km	4,2 km	3,9 km	4,4 km	4.4 km
Durée	moy	15.9 min	16,28 min	16,9 min	19,2 min	15,1 min	
	C.V	181%	111%	124%	124%	178%	
	med	12,6 min	13,4 min	12,9 min	15,0 min	12,2 min	
COURSES ANNULÉES							
	Nombre de courses annulées par jour						
	Lundi					285 courses	83%
	Mardi					301 courses	63%
	Mercredi					324 courses	63%
	Jeudi					521 courses	122%
	Vendredi					534 courses	106%
	Samedi					528 courses	110%
	Dimanche					320 courses	95%
USAGER	Temps d'attente pour une course réalisée					9,3 min	63%
	Nombre d'usagers actifs par jour					1 018 usagers	47%
	Nombre de courses réalisées par usager actif, par jour					1,19 courses	4%
	Nombre de courses annulées par usager actif, par jour					1.37 courses	9%

Durant la période été-automne, le nombre de courses réalisées par toute l'équipe de chauffeurs Téo augmente tous les jours de la semaine pour atteindre son pic avec plus de 1250 courses le vendredi. Le nombre de courses diminue en fin de semaine pour atteindre un peu plus de 645 courses le dimanche. Le coefficient de variation se trouve entre 35-40 %. Rappelons que l'offre de Téo a constamment augmenté et que la demande a suivi durant la période été-automne. Si la période d'automne seule est considérée, le nombre moyen de courses par jour double, voire triple.

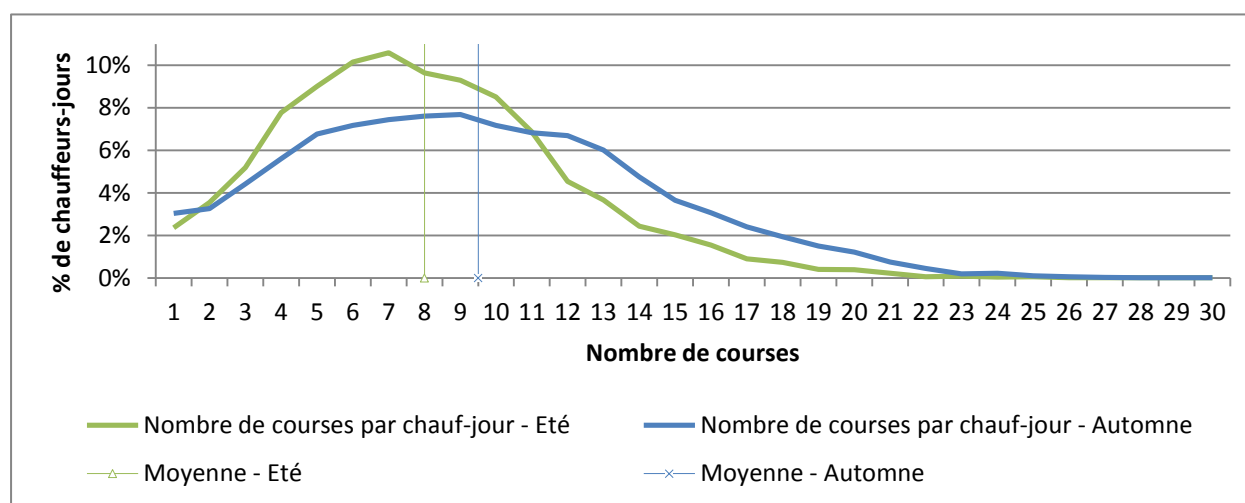


Figure 4-8 : Distribution des chauffeurs selon le nombre de courses effectuées par jour

Un chauffeur réalise près de 9 courses par jour. La figure 4-8 montre la distribution des chauffeurs en fonction du nombre de courses que chacun réalise chaque jour de l'été ou de l'automne. Faire 4 à 12 courses par jour est donc habituel pour un chauffeur de taxi. Le graphique nous montre que certains vont jusqu'à 22 courses par jour.

Lorsque nous classons les chauffeurs dans l'ordre décroissant du nombre de courses (respectivement du nombre de revenu généré) qu'ils ont faites durant chacune des saisons, nous obtenons la figure 4-9 (respect. la figure 4-10). Que ce soit en été ou en automne, 50% des courses ont été réalisées par 30% des chauffeurs et 50% des chauffeurs ont fait de l'ordre de 74% des courses. En ce qui concerne les revenus générés, 50% des revenus sont générés par 29% des chauffeurs l'été et l'automne, et 50% des chauffeurs génèrent 76% des revenus en été et 74% en automne.

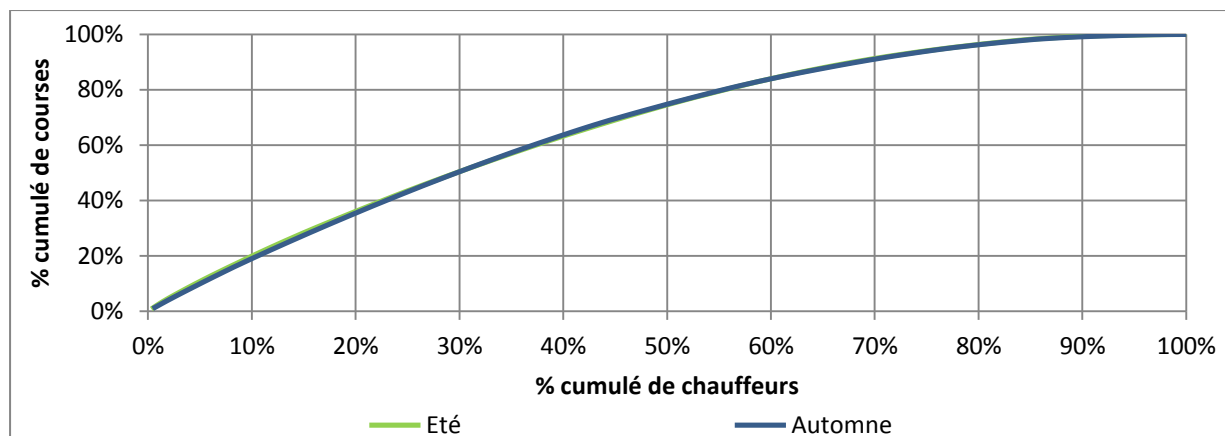


Figure 4-9 : Distribution cumulée du nombre de courses en fonction du pourcentage cumulé des chauffeurs

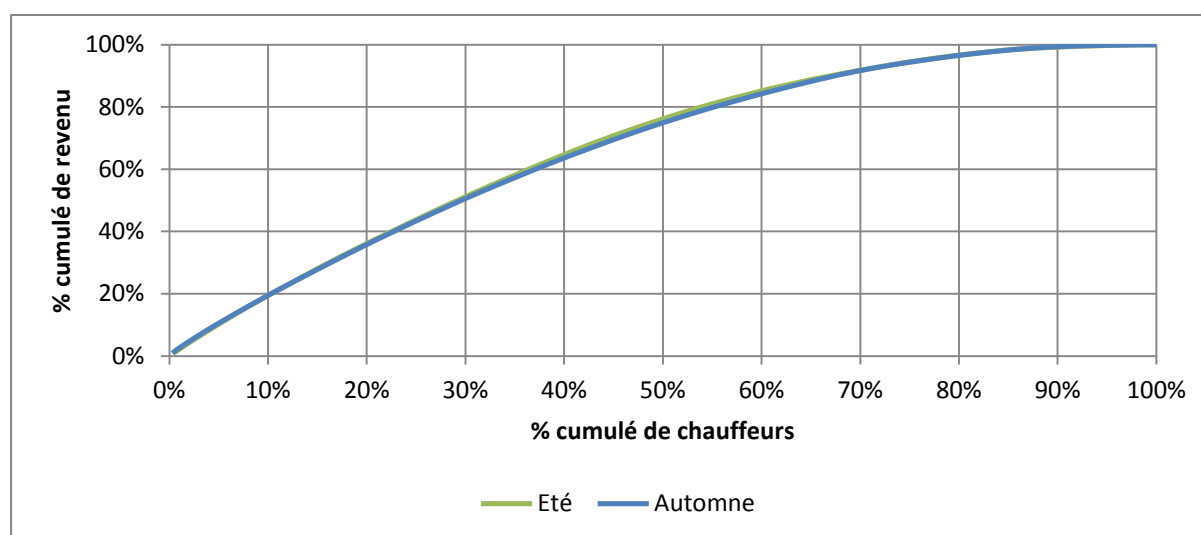


Figure 4-10 : Distribution cumulée du revenu généré en fonction du pourcentage cumulé des chauffeurs

La figure 4-11 et la figure 4-12 donnent l'évolution du pourcentage de courses réalisées par heure-chauffeur travaillée (en rouge) chaque semaine pour l'été et l'automne. L'analyse ne se fait donc pas à l'échelle du chauffeur, mais à l'échelle de l'équipe de chauffeurs Téo. Une heure-chauffeur travaillée équivaut à 1 h de travail de toute l'équipe. Un léger motif se dégage de semaine en semaine pour les deux saisons. Nous pouvons observer l'apparition d'un pic en milieu de semaine (de mercredi à samedi) qui diminue en fin de semaine. Globalement, chaque semaine, le pourcentage de courses par heure-chauffeur travaillée est toujours au-dessus de 50%. Donc chaque semaine, plus d'une course est satisfaite à chaque 2 heures-chauffeurs

travaillée. Le nombre d'heures-chauffeurs travaillées est aussi donné (en vert) et suit l'augmentation du nombre de courses par heure-chauffeur travaillée. Ainsi plus il y a de courses, plus le nombre d'heures travaillées augmente. Cependant une forte augmentation du nombre de courses n'entraîne pas nécessairement une forte augmentation du nombre d'heures-travaillées comme le montre la troisième semaine du mois de juillet.

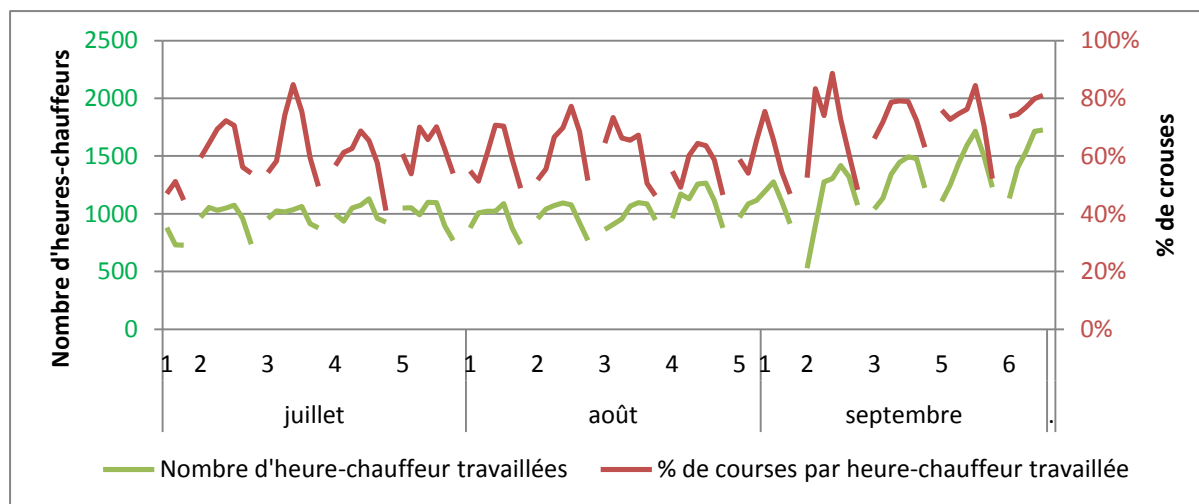


Figure 4-11 : Évolution du nombre de courses par heure-chauffeur travaillée - Été

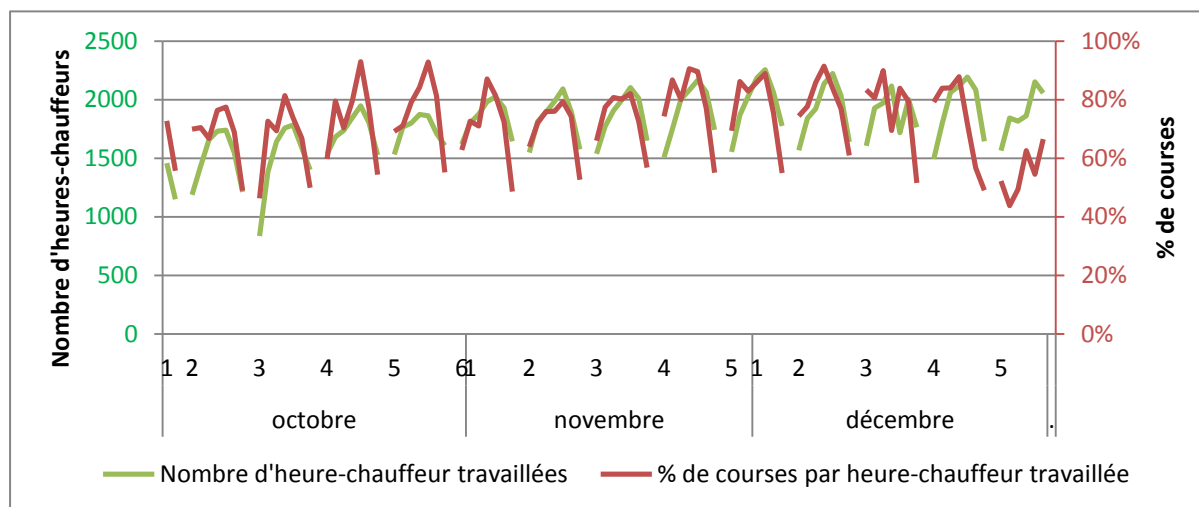


Figure 4-12 : Évolution du nombre de courses par heure-chauffeur travaillée - Automne

Entre l'été et l'automne, le nombre d'heures-chauffeurs travaillé a augmenté puisque le nombre de chauffeurs lui-même a augmenté. De plus, contrairement à l'été, le pourcentage de courses par heure-chauffeur travaillée tend chaque semaine d'automne à dépasser les 80%.

En ce qui concerne les courses, leur longueur moyenne est de 6,3 km (médiane : 4,4 km). Les

figures ci-dessous précisent notre analyse. La figure 4-13 montre le profil des longueurs et des vitesses de courses pour une semaine moyenne. Ce profil est quasiment le même entre l'été et l'automne. Un motif se dégage : chaque jour entre 4-6 h du matin les courses sont les plus longues. Leurs longueurs dépassent les 12 km le lundi, diminuent jusqu'à atteindre les 10 km le vendredi et remontent petit à petit jusqu'à 11 km le dimanche. Après 6 h, les longueurs de courses tournent autour de 6 km sauf le dimanche où l'après-midi elles peuvent dépasser 8 km. Quand les courses sont les plus longues, entre 0-6 h du matin, les taxis roulent le plus vite. Leur vitesse peut atteindre les 35 km.h⁻¹. Entre 6-18 h, les taxis gardent une vitesse moyenne autour de 20 km.h⁻¹. Et après 18 h, leur vitesse réaugmente même si les longueurs de courses, elles, n'augmentent pas. Le trafic routier entre 6-18 h a probablement un impact sur la vitesse de courses des taxis.

L'expansion de la couverture du réseau de Téo n'a pas eu d'impact sur le profil des longueurs et des vitesses de courses. La seule différence notoire est qu'à l'automne les courses sont plus longues qu'en été. Les vitesses de courses ne sont pas plus élevées.

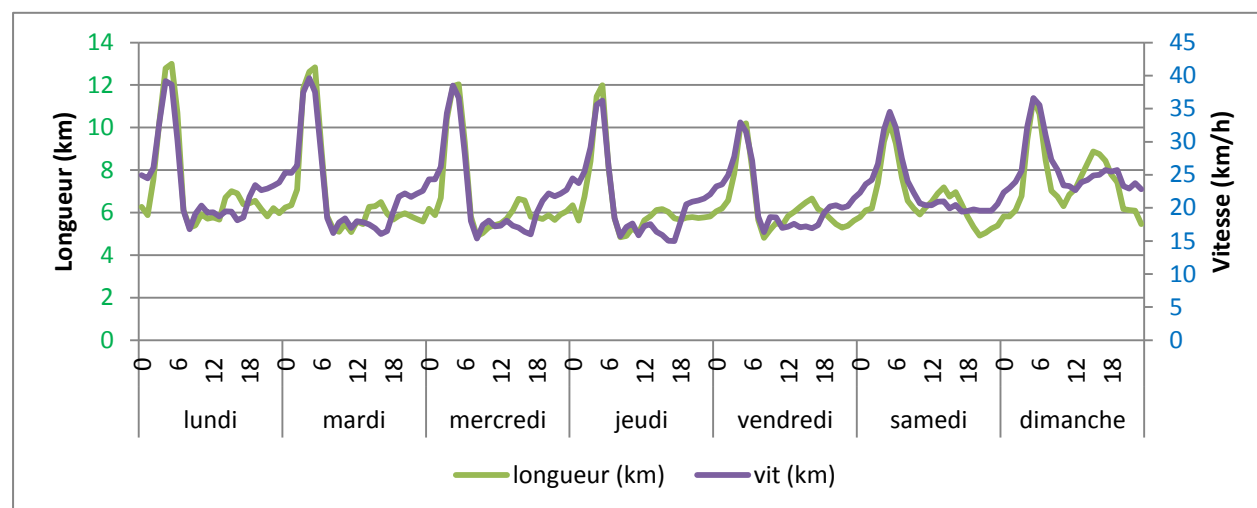


Figure 4-13 : Comparaison des longueurs et vitesses moyennes pour une semaine moyenne (Été + Automne)

D'après le tableau 4-3, les courses entre 0-6h du matin font près de 9km en moyenne, puis leurs longueurs tournent autour de 6km pour les périodes horaires suivantes. Par période, la figure 4-14 et figure 4-15 montrent la distribution du pourcentage de courses réalisées en fonction de leurs longueurs. Les courses faisant 3-5 km sont majoritaires entre 0-6h, et celles faisant 2-4 km sont majoritaires pour les autres périodes horaires de la journée. Ces tendances

restent valables aussi bien à l'été qu'à l'automne. La seule différence réside dans les courses les plus longues : en été, un pic à 19 km est marqué tandis que ce pic arrive à 20 km pour l'automne.

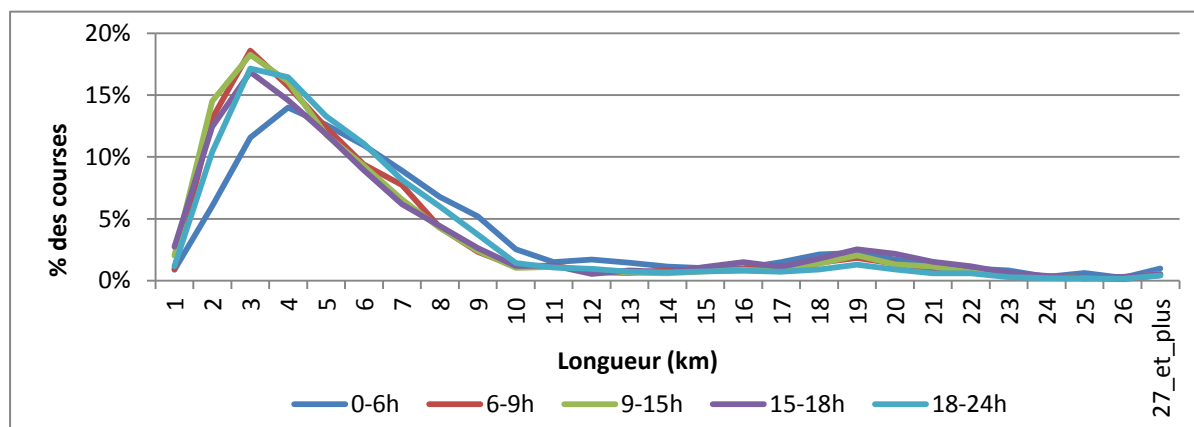


Figure 4-14 : Distribution des longueurs de courses par période de la journée - Été

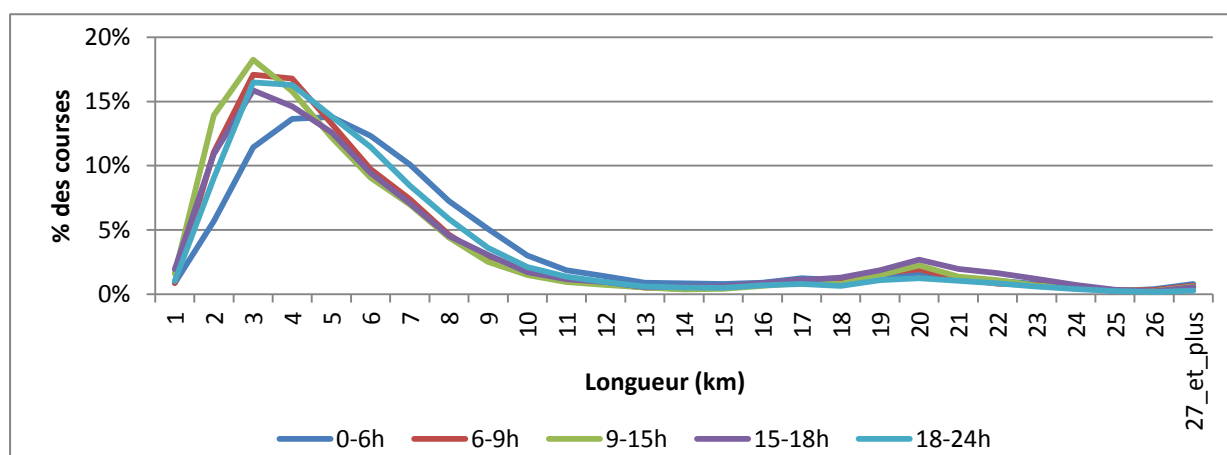


Figure 4-15: Distribution des longueurs de courses par période de la journée - Automne

En ce qui concerne les durées de courses, leur durée moyenne dépasse les 16 min (médiane : 13 min). Entre 0 et 6 h du matin, même si les courses sont les plus longues de la journée, leur durée moyenne est autour de 16 min comme le reste de la journée, la période 15 et 18 h mise à part. En effet durant cette période la durée moyenne d'une course est de plus de 19 min. Étant une période de pointe, le trafic routier augmente probablement la durée des courses. Les figure 4-16 et figure 4-17 corroborent notre analyse. La course ayant eu lieu entre 15 et 18 h tendent à durer plus longtemps que les autres périodes.

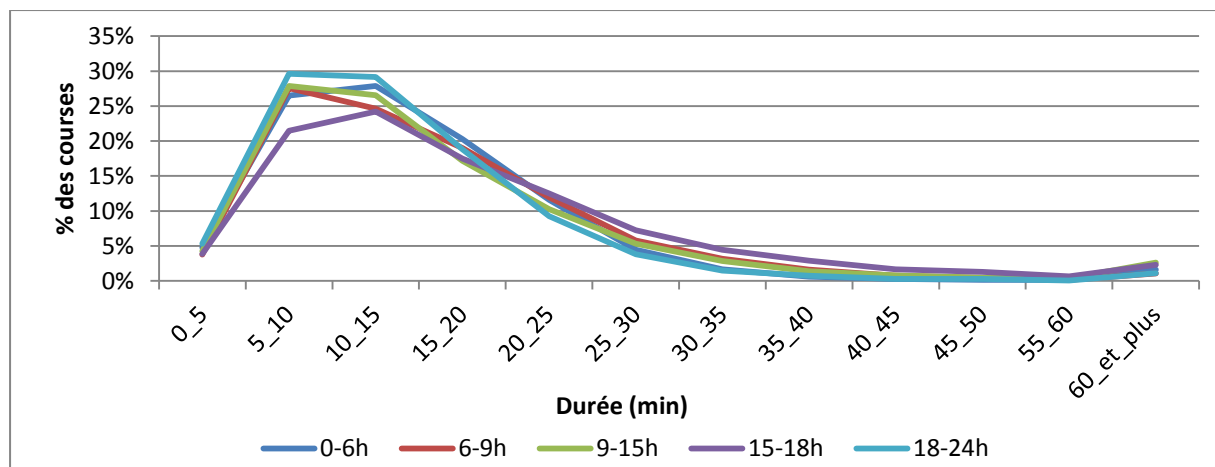


Figure 4-16 : Distribution des durées des courses par période de la journée - Été

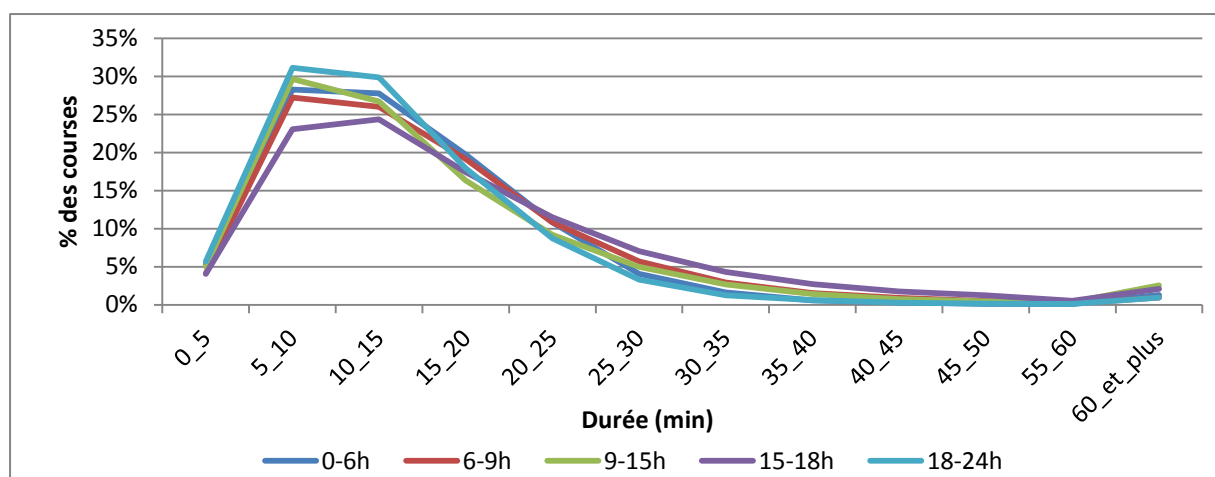


Figure 4-17 : Distribution des durées de courses par période de la journée - Automne

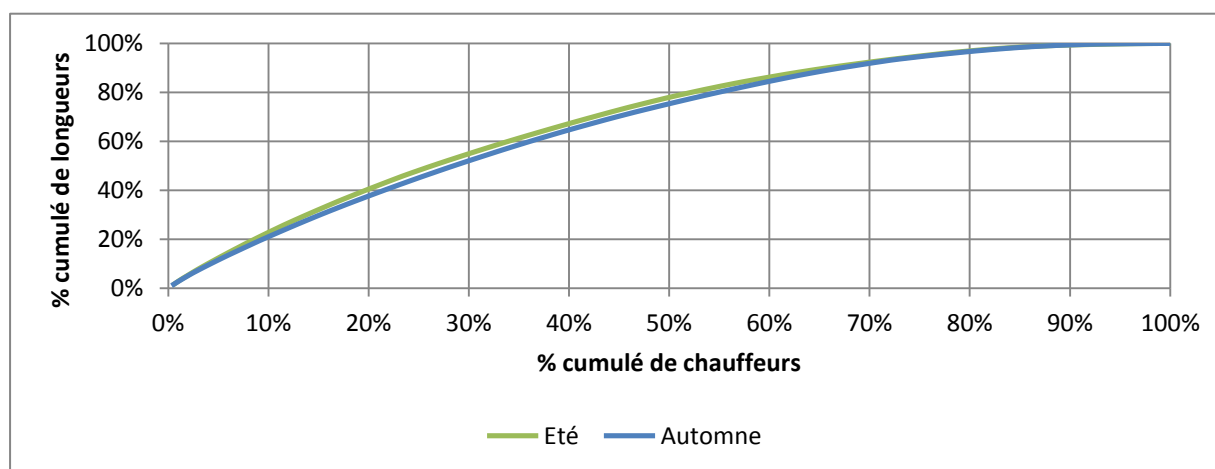


Figure 4-18 : Distribution cumulée des longueurs des courses en fonction du pourcentage cumulé des chauffeurs.

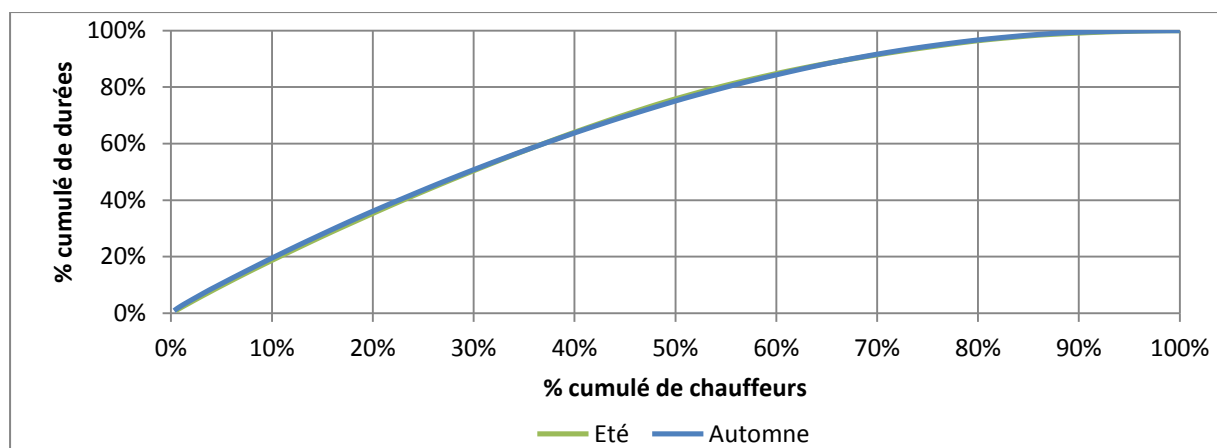
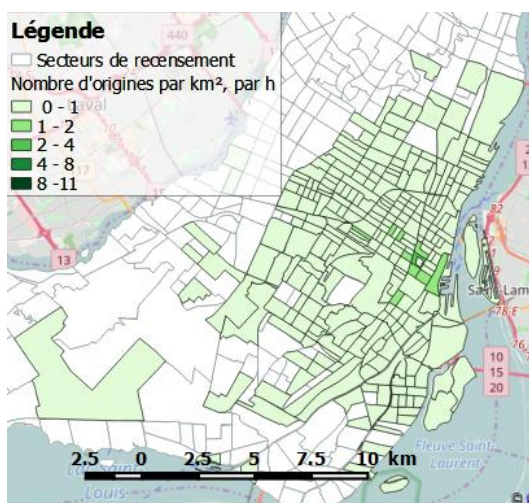


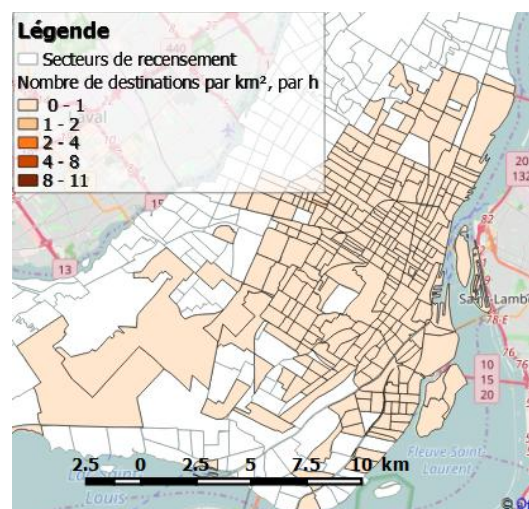
Figure 4-19 : Distribution cumulée des durées de courses en fonction du pourcentage cumulé des chauffeurs

Pour finir l'analyse des durées et longueurs de courses, les figure 4-18 et figure 4-19 donnent le pourcentage cumulé de durée et de longueur de courses en fonction du pourcentage cumulé de chauffeurs. Ainsi, nous observons que 50% du total des longueurs de courses a été fait par 26% des chauffeurs en été, 28% en automne. Et que 50% du total des durées de courses a été fait par 30% des chauffeurs en été, 29% en automne.

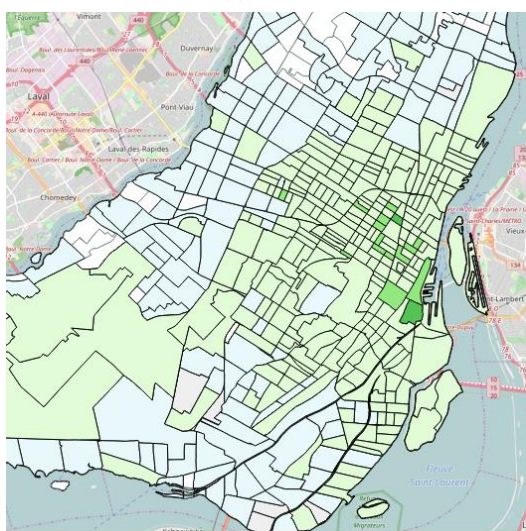
La figure 4-0 (respect. figure 4-1) donne la densité des origines et des destinations des courses par secteurs de recensement pour une journée moyenne de la saison d'été (respect. d'automne). En été, entre 6 et 18 h, la densité d'origines est moins concentrée que la densité de destinations puisqu'il y a entre 0 et 4 origines par km² contre 0 et 8 destinations par km² selon la période de la journée. Les origines et les destinations des courses sont plus denses en centre-ville et autour du Vieux-Port. Si leur densité n'est pas très forte de 0 à 6h du matin, elle augmente progressivement durant la journée. Les origines proviennent majoritairement du centre-ville et du Plateau Mont-Royal durant toutes les périodes. Mais les destinations sont majoritairement au centre-ville de 6 à 15 h, puis se répartissent entre le centre-ville et le Plateau de 15 à 24 h. Ces tendances restent vraies à l'automne et même s'intensifient. La densité d'origines et de destinations augmente notamment vers Rosemont-Petite-Patrie, passant de 0-1 à 2-4 origines et destinations par km², par h suivant les périodes de la journée.



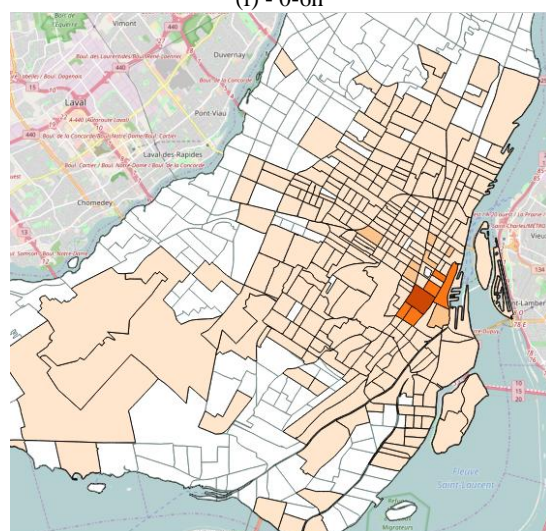
(a) - 0-6h



(f) - 0-6h



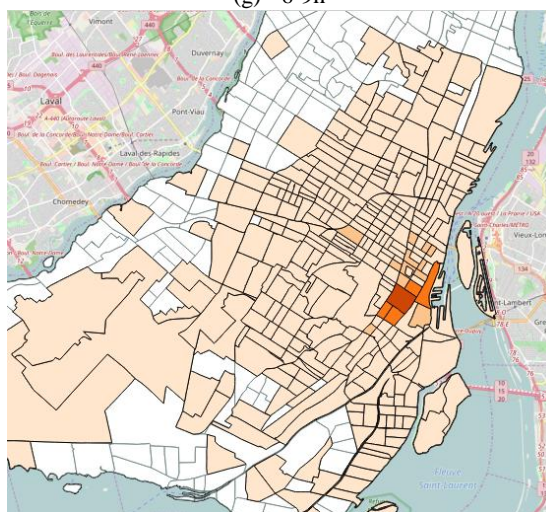
(b) - 6-9h



(g) - 6-9h

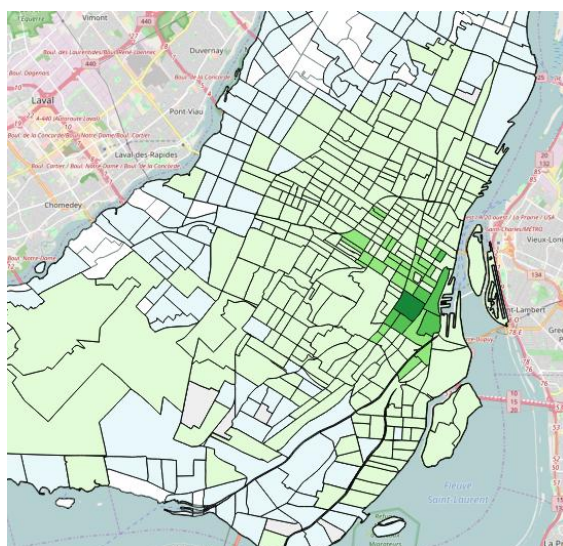


(c) - 9-15h

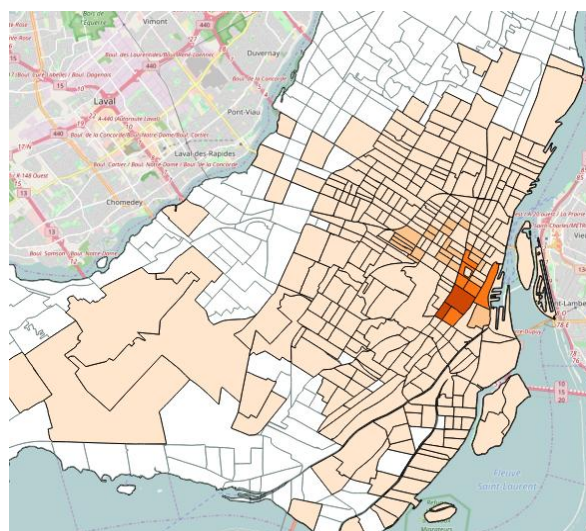


(h) - 9-15h

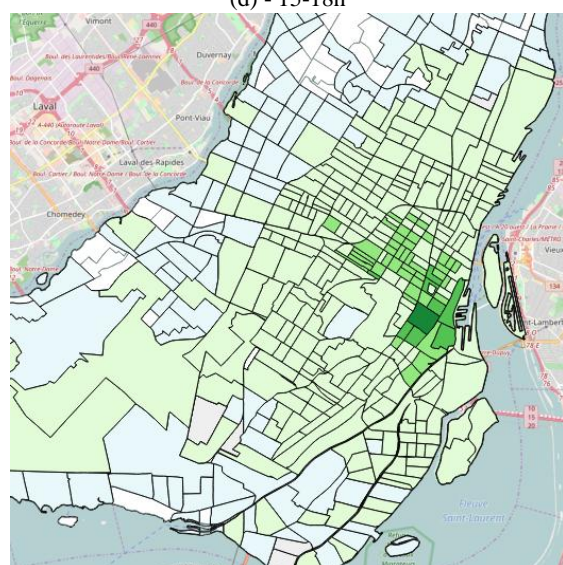
Figure 4-20 : Densité horaire d'origines et de destinations par secteur de recensement pour une journée moyenne (tous types de services confondus). - Été



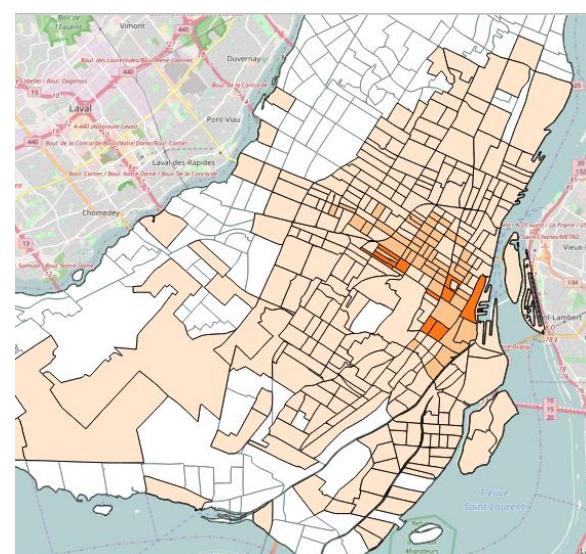
(d) - 15-18h



(i) - 15-18h

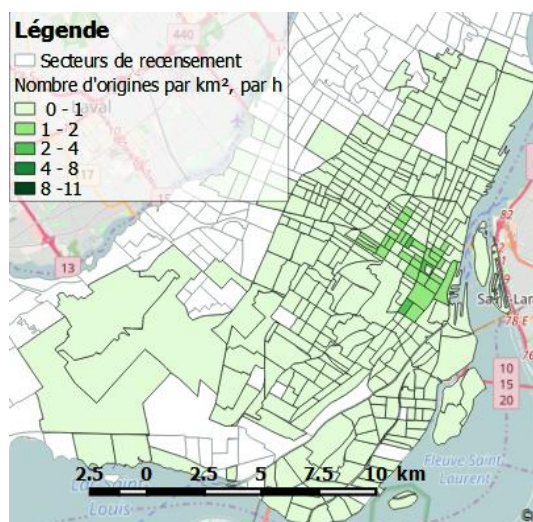


(e) - 18-24h

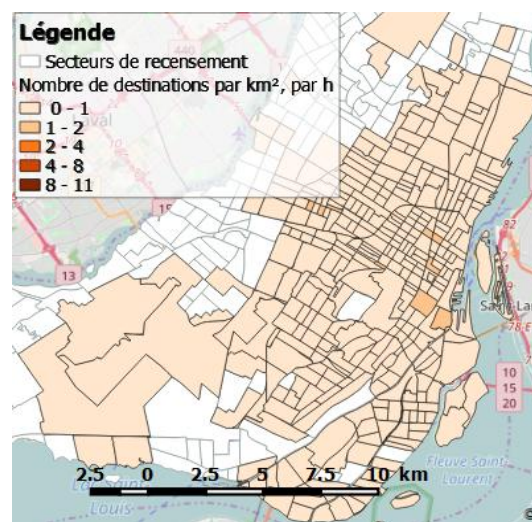


(j) - 18-24h

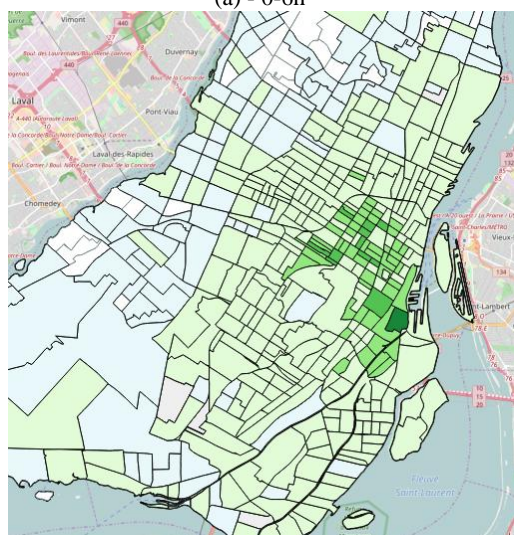
Figure 4-20 : Densité horaire d'origines et de destinations par secteur de recensement pour une journée moyenne (tous types de services confondus). - Été (suite)



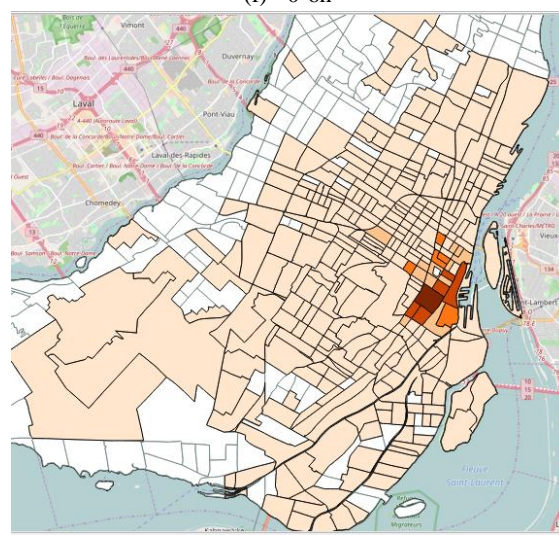
(a) - 0-6h



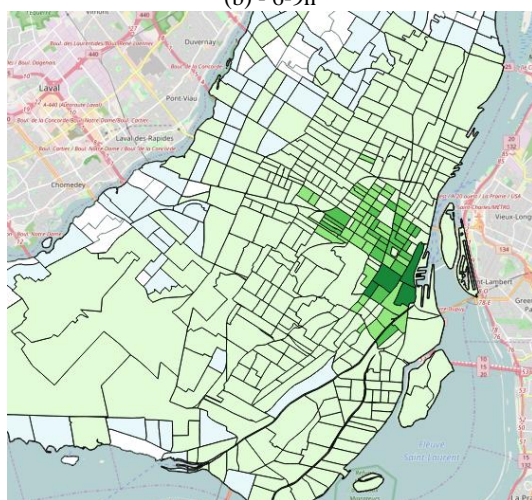
(f) - 0-6h



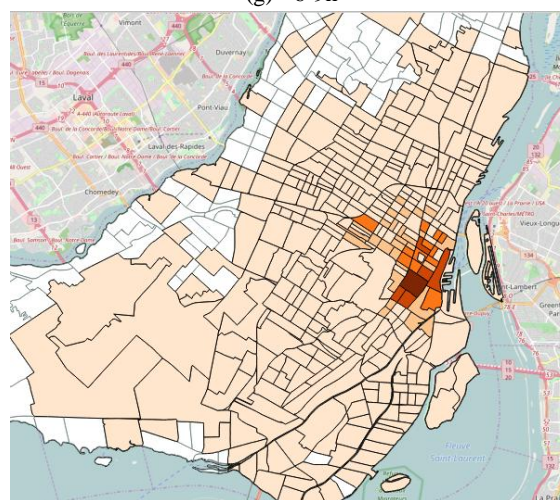
(b) - 6-9h



(g) - 6-9h



(c) - 9-15h



(h) - 9-15h

Figure 4-21 : Densité horaire d'origines et de destinations par secteur de recensement pour une journée moyenne (tous types de services confondus). - Automne

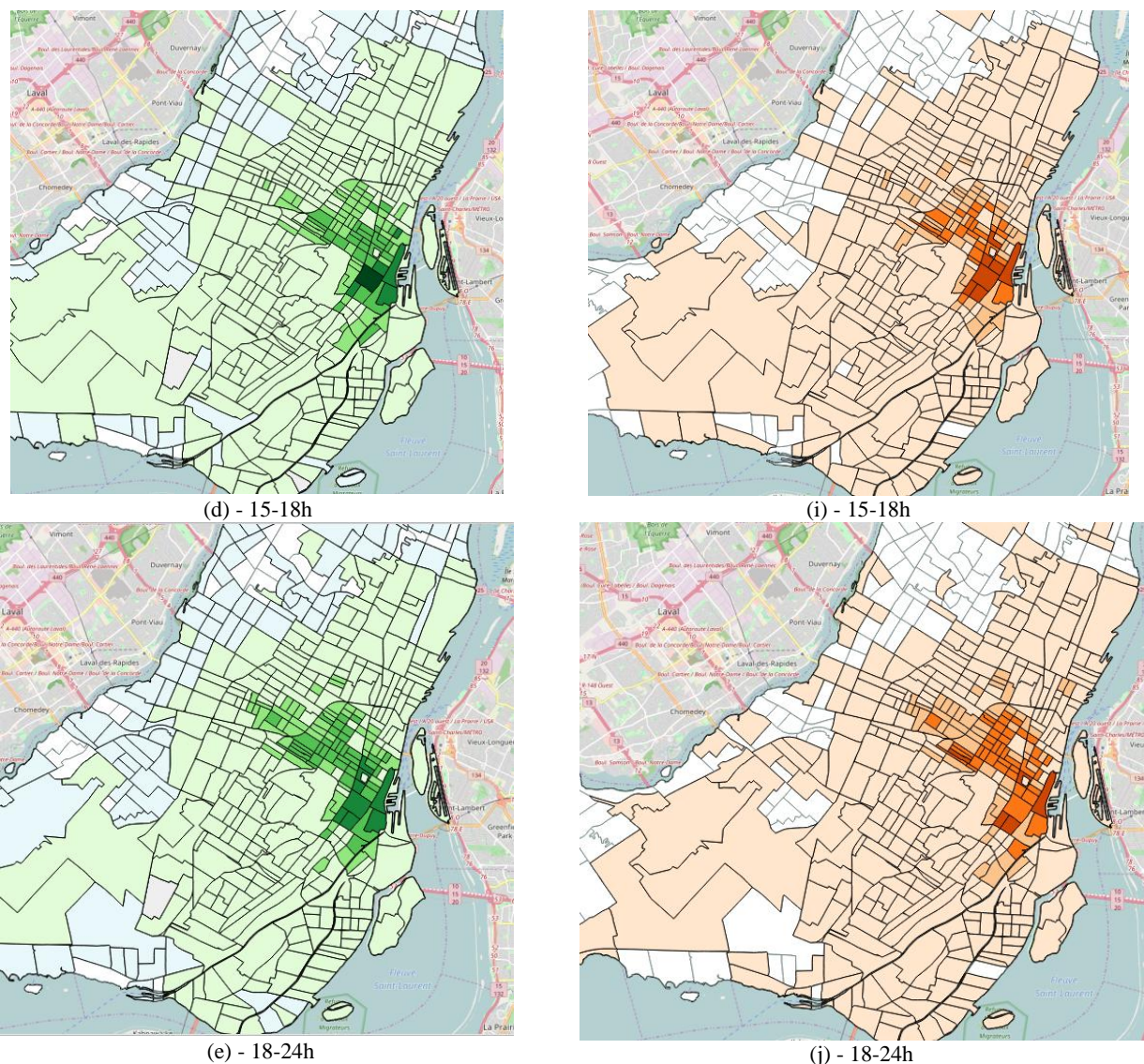


Figure 4-21 : Densité horaire d'origines et de destinations par secteur de recensement pour une journée moyenne (tous types de services confondus). - Automne (suite)

En ce qui concerne les modes de réservations, il existe à ce jour 4 façons différentes de réserver sa course : par application mobile, par téléphone, par hélage et par site internet. L'application mobile était la seule manière de commander une course jusqu'à l'autorisation des courses par hélage pour les usagers particuliers. Les courses hélées commencent le 09/11/2016 et sont au nombre de 90 entre cette date et le 31/12/2016. Les usagers corporatifs pouvaient directement commander leurs courses par téléphone. Les données montrent qu'aucun usager corporatif n'a commandé de courses par application mobile, néanmoins il se peut qu'ils l'aient fait en se présentant comme usager particulier. À partir du 08/11/2016, certaines compagnies pouvaient

commander leurs courses par le portail (le site internet) de Téo Taxi. Le nombre de courses par internet était au nombre de 216 entre le 08/11/2016 et le 31/12/2016.

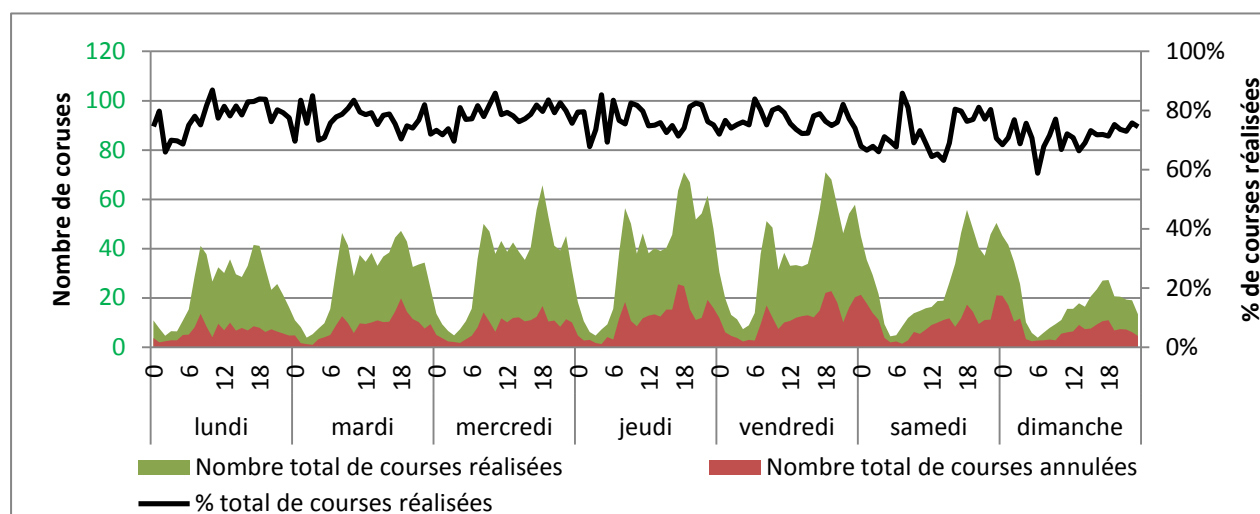


Figure 4-21 : Taux de courses réalisées - Eté

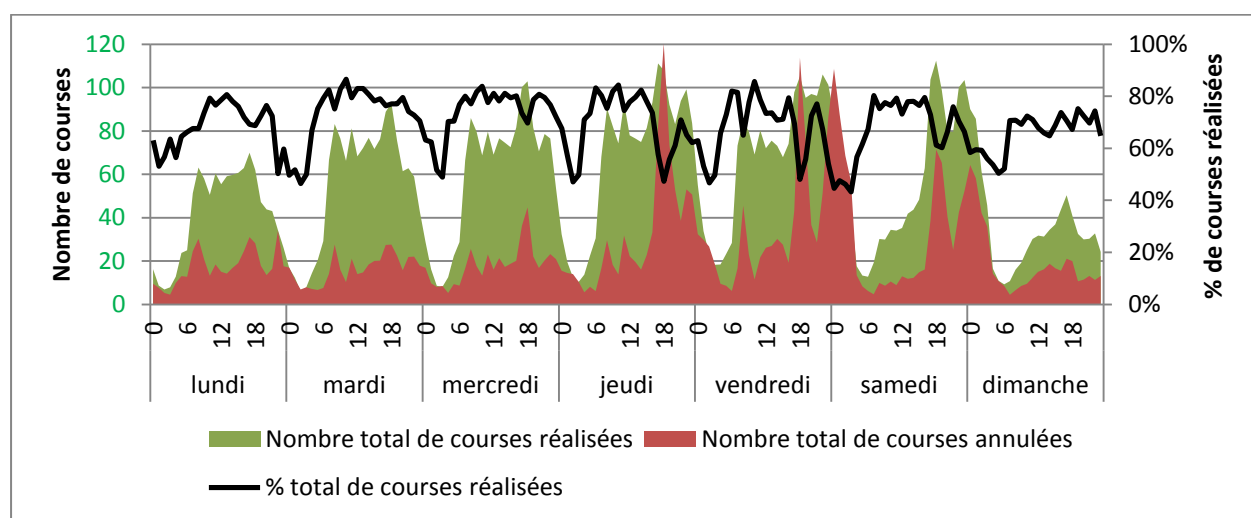


Figure 4-22 : Taux de courses réalisées - Automne

Entre l'été et l'automne, le nombre de courses réalisées d'une part et le nombre de courses annulées d'autre part ont quasiment doublé chaque jour (figure 4-21 et figure 4-22). De mardi à vendredi, de 6 à 18h, le nombre de courses réalisées est entre 20 et 40 par heure l'été et entre 60 et 80 par heure l'automne. Pour ces mêmes jours et ces mêmes horaires, le nombre de courses annulées est entre 10 et 20 par heure l'été. En automne, il est autour de 20 du lundi au mercredi.

Mais du jeudi au samedi, Téo Taxi connaît une explosion des annulations en fin d'après-midi à partir de 17 h.

Le profil du taux de courses réalisées varie considérablement entre l'été et l'automne. Après une analyse plus approfondie, le profil pour les mois d'octobre et de novembre ressemble à celui d'été. En été le taux de courses réalisées dépasse globalement les 70%. 7 courses sur 10 sont donc réalisées chaque heure de la semaine. C'est l'activité des taxis durant le mois de décembre qui impacte le taux de courses réalisées pour toute la période d'automne. Durant ce mois, le taux est très irrégulier. Durant certaines heures de certaines journées, il est au-dessous de 30%. Le froid de l'hiver québécois est une cause probable de perturbation de la technologie électrique des véhicules provoquant de ce fait des pics d'annulations de courses.

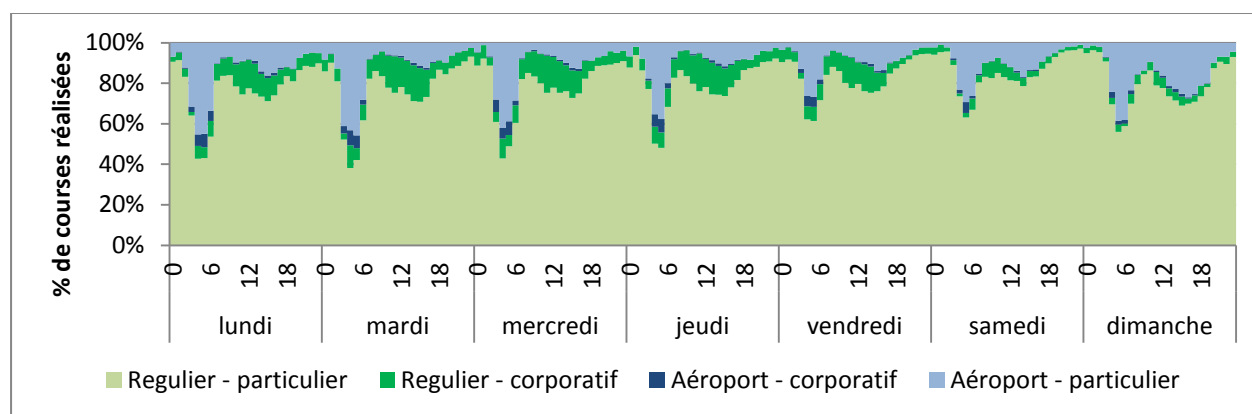


Figure 4-23 : Pourcentage de courses réalisées par types de services et de clients pour une semaine moyenne - (Été + Automne)

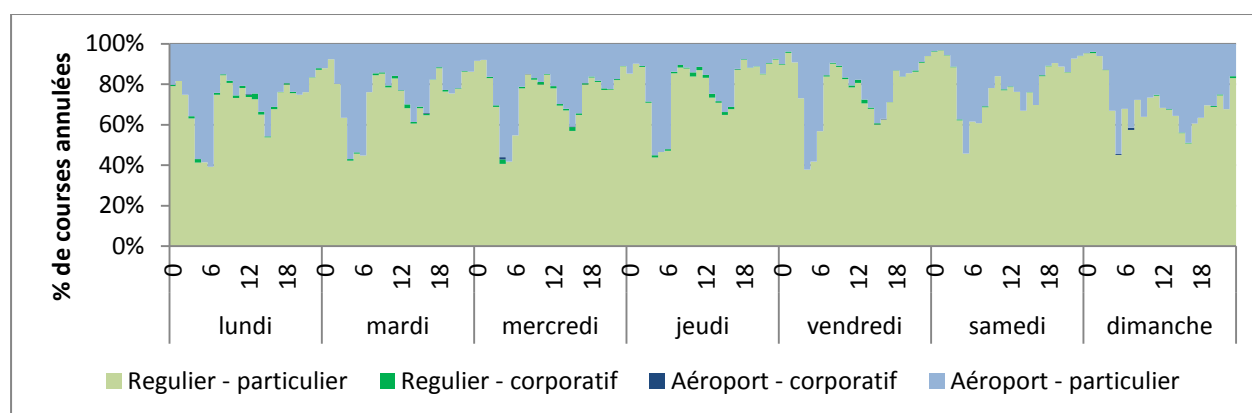
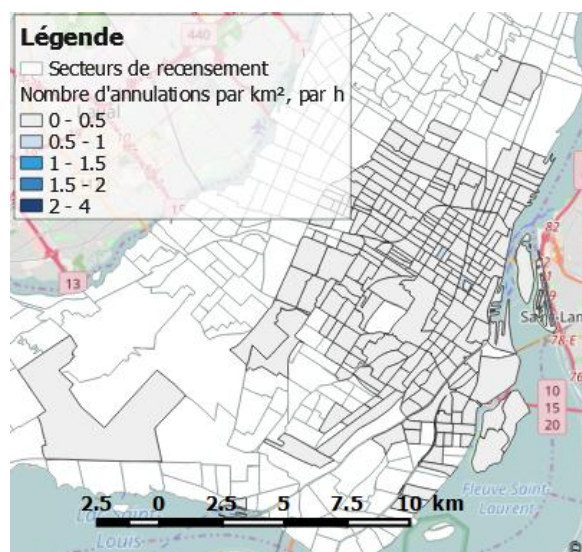


Figure 4-24 : Pourcentage de courses annulées par types de services et de clients pour une semaine moyenne - (Été + Automne)

Rappelons que Téo Taxi offre deux types de services à deux types d'usagers. Cette compagnie permet aux usagers particuliers et aux usagers corporatifs de réaliser des courses originant ou se destinant à l'aéroport, ce sont les courses aéroportuaires, mais aussi à toute origine et destination de leur choix, ce sont les courses régulières.

La Figure 4-23 montre que la majorité des courses réalisées servent des usagers particuliers pour des courses régulières. Pour tous les jours de la semaine et à toute heure de la journée, le pourcentage de ce type de courses tournent autour de 80% du total de courses réalisées sauf entre 3 et 7 h du matin où les courses aéroportuaires pour les usagers particuliers montent à plus 60%. Il est notable aussi que les courses régulières d'usagers corporatifs se fassent principalement entre 8 et 15 h pour une journée ouvrable. La figure 4-24 montre que les courses annulées sont majoritairement issues des usagers particuliers que ce soit pour la réalisation d'une course aéroportuaire ou d'une course régulière. Autour de 80% des courses annulées sont annulées par les usagers particuliers pour le service régulier et autour de 20% par ces mêmes usagers pour le service aéroportuaire, à toute heure de la journée sauf entre 3-7 h du matin. En effet durant cette période, ce sont les courses aéroportuaires qui composent autour de 60% du total des courses annulées et les courses régulières font près des 40% restant. Ces profils sont similaires pour l'été et l'automne.

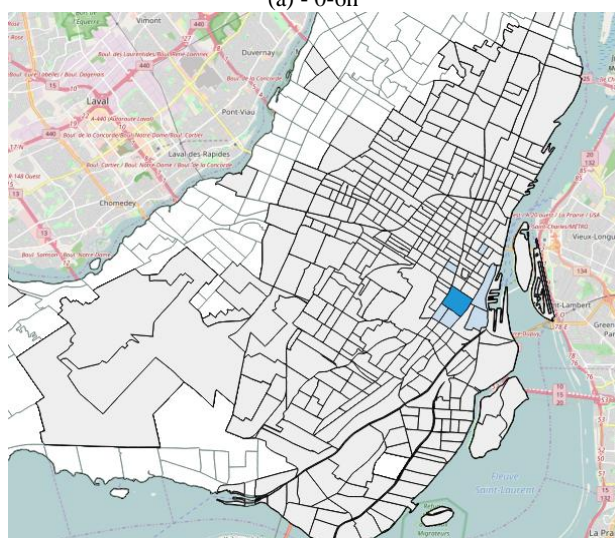
L'analyse temporelle des courses annulées est complétée avec une analyse spatiale. La figure 4-25 donne pour chaque période d'une journée moyenne d'été, la densité d'annulations par secteur de recensement. Cette densité peut atteindre les 2-4 courses par km², par heure à partir de 9 h en centre-ville. La densité d'annulations est plus intense à l'automne. Alors qu'à l'été il n'y avait quasiment pas d'annulations sur le Plateau Mont-Royal, en automne il y en a jusqu'à 2 par km² par h entre 0-6 h du matin et 18-24 h.



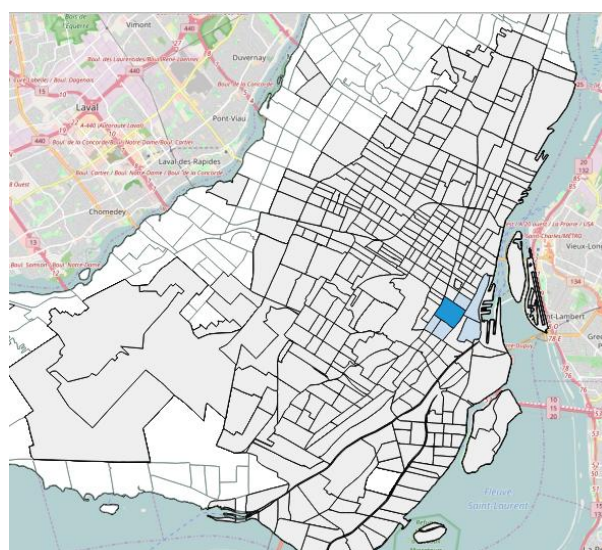
(a) - 0-6h



(b) - 6-9h



(c) - 9-15h

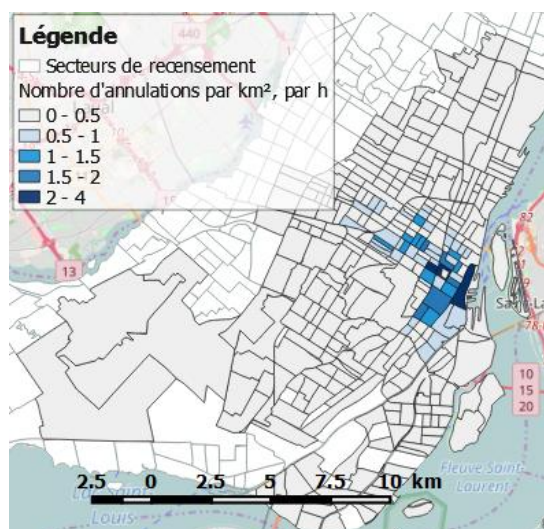


(d) - 15-18h



(e) - 18-24h

Figure 4-25 : Densité horaire d'annulations par secteur de recensement pour une journée moyenne (tous types de services confondus) entre (a) 0-6h, (b) 6-9h, (c) 9-15h, (d) 15-18h, (e) 18-24h. - Été



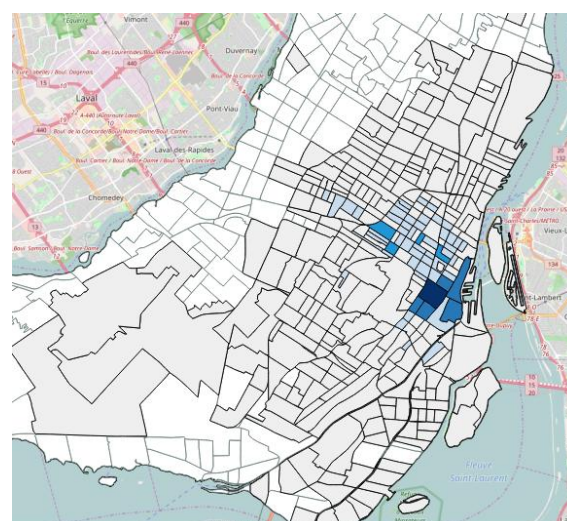
(a) - 0-6h



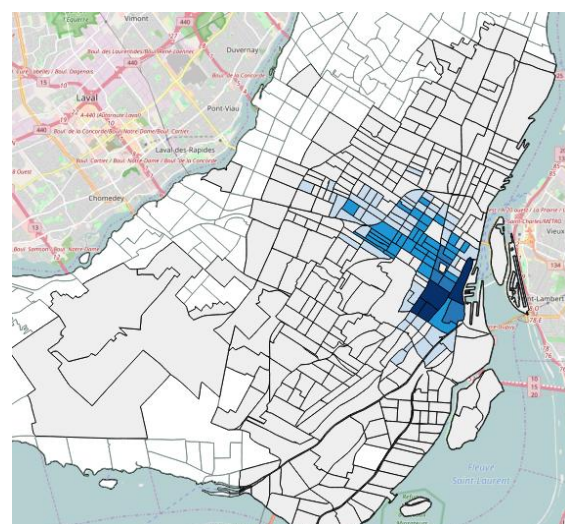
(b) - 6-9h



(c) - 9-15h



(d) - 15-18h



(e) - 18-24h

Figure 4-26 : Densité horaire d'annulations par secteur de recensement pour une journée moyenne (tous types de services confondus) entre (a) 0-6h, (b) 6-9h, (c) 9-15h, (d) 15-18h, (e) 18-24h. - Automne

Le temps d'annulation d'une course est le temps que met un usager pour annuler sa course après l'avoir commandée. La figure 4-27 donne le nombre cumulé de courses annulées en fonction du temps d'annulation. Parmi l'ensemble des courses annulées, plus de 40% ont été annulées dans un intervalle de temps inférieur à 1 min et 20% dans un intervalle compris entre 1 et 2 min. En conclusion un peu plus de 60% des courses ayant été annulées l'ont été dans intervalle de temps inférieur à 2 min. La rapidité du temps d'annulation donne une indication sur la temps que prend un usager avant de décider d'annuler sa course. Certains d'entre eux peuvent donc attendre au-delà de 5 min.

La raison pour laquelle de nombreuses courses sont très vite annulées n'est pas identifiée. Il est possible que l'utilisateur soit rapidement informé par l'application mobile de la disponibilité de son véhicule et qu'il puisse, dès lors, décider rapidement de l'annulation de sa course en cas de non-satisfaction du temps d'attente annoncé. Bien que la base de données de Téo Taxi contienne les motifs des annulations, l'étude de ces dernières n'est pas permise car la validité des raisons données par les usagers est questionable. En outre, la compagnie a modifié à différentes reprises les règles d'acceptation des courses et les motifs possibles d'annulation ce qui rend difficile le suivi longitudinal de ce phénomène.

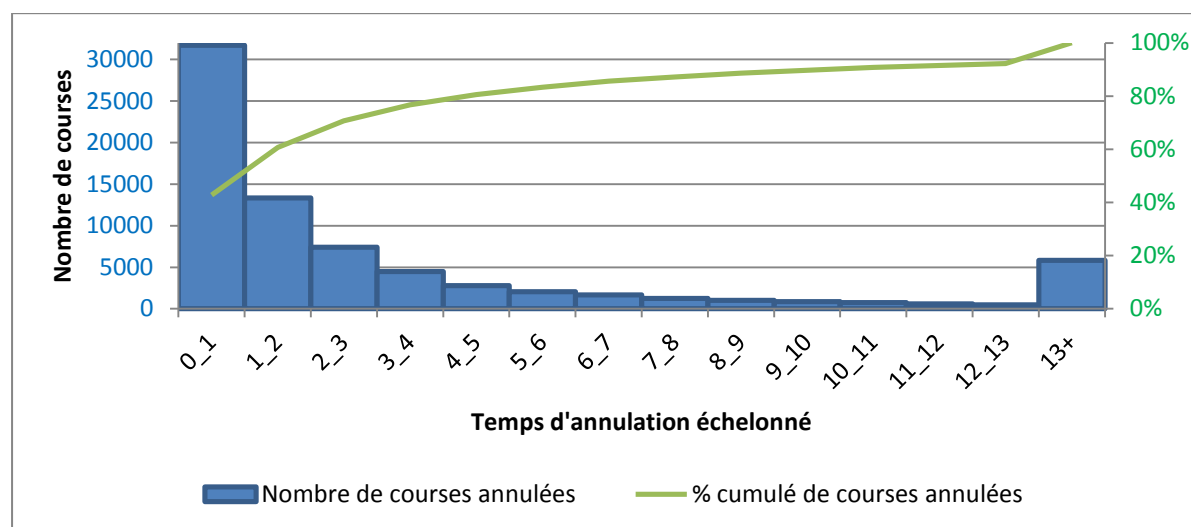


Figure 4-27 : Nombre cumulé de courses annulées en fonction du temps d'annulation - (Été + Automne)

La figure 4-28 montre la distribution du nombre d'utilisateurs par nombre d'annulations pour les deux saisons. Les types d'annulations pris en compte ne concernent que ceux dont le client est

responsable, c'est-à-dire toutes sauf la raison d'annulations "autre". Sur les 45 128 usagers uniques de la période juillet-décembre 2016, 27% (12 000 usagers) ont annulé au moins une fois leur course et moins de 13% l'ont au moins annulée deux fois. Parmi l'ensemble des usagers ayant annulé au moins une fois leurs courses, 50% l'ont annulée une seule fois et 10% l'ont annulée 4 fois.

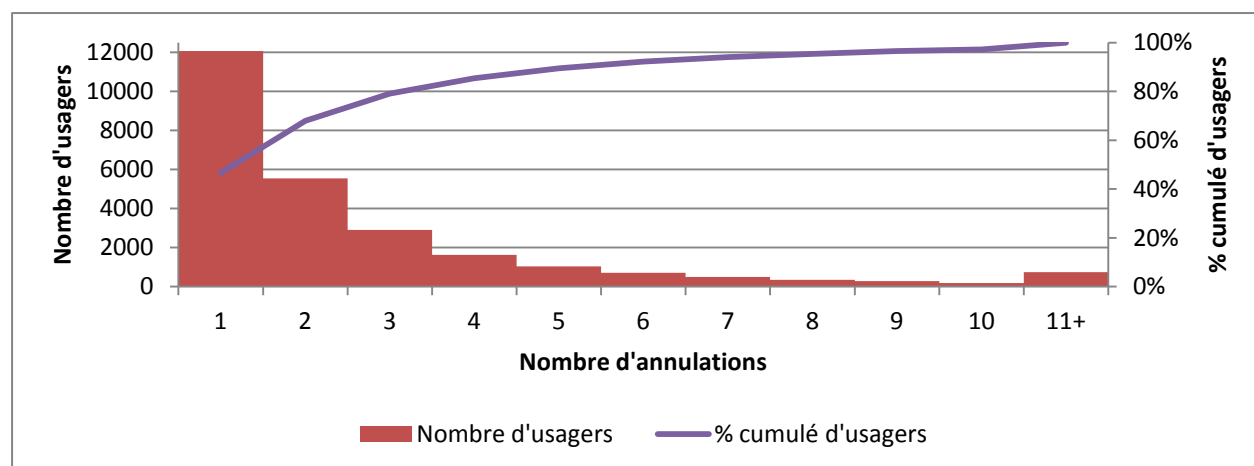


Figure 4-28 : Distribution du nombre d'annulations par nombre d'usagers - (Été + Automne)

Quand un usager n'annule pas sa course, les figure 4-29 et figure 4-30 donnent la longueur moyenne de cette dernière en fonction du temps d'attente de l'usager. Ce graphique indique combien de temps un usager est prêt à attendre pour faire une course de longueur variable. Il est remarquable que pour les deux saisons, plus la course est longue plus l'usager est prêt à attendre. Cependant après 15 min d'attente, les courses réalisées sont de moins en moins nombreuses. Un usager faisant une longue course est peut-être moins bien desservi par les différentes offres de transports de sa ville, il peut dès lors être plus enclin à attendre plus longtemps son taxi. Le pic à 20 min de la figure 4-30 est difficilement explicable. Le nombre de courses est très faible, moins de 1000 pour tout l'automne, près de la moitié ont eu lieu en décembre 2016.

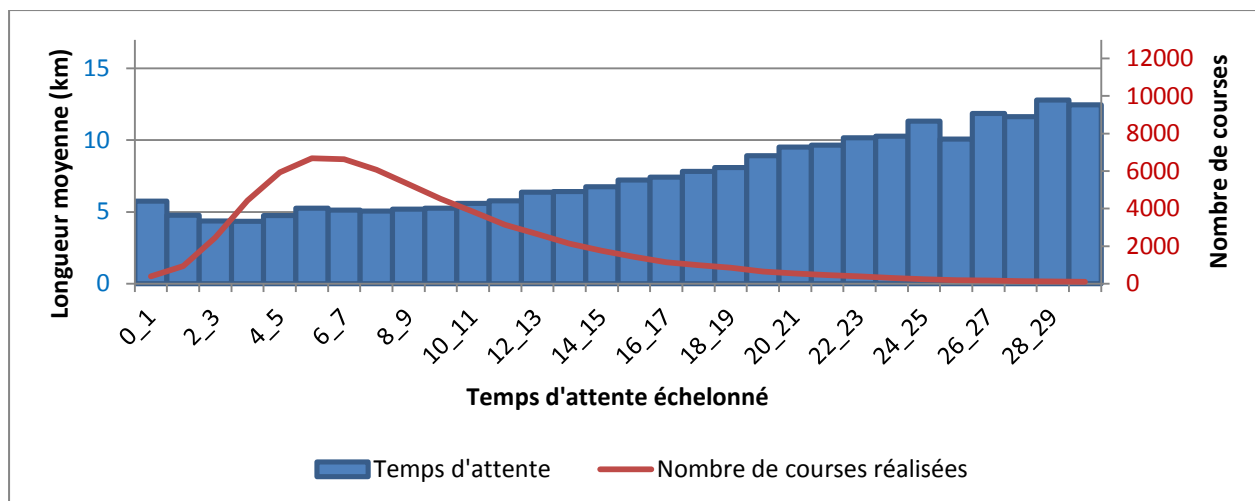


Figure 4-29 : Longueur moyenne d'une course réalisée selon le temps d'attente de l'utilisateur - Été

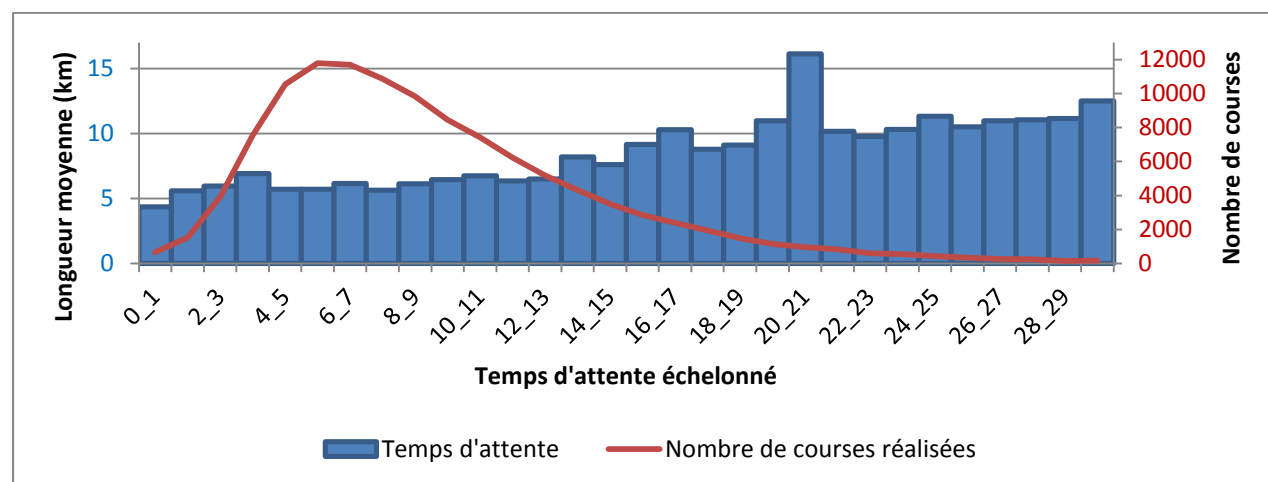


Figure 4-30 : Longueur moyenne d'une course réalisée selon le temps d'attente de l'utilisateur - Automne

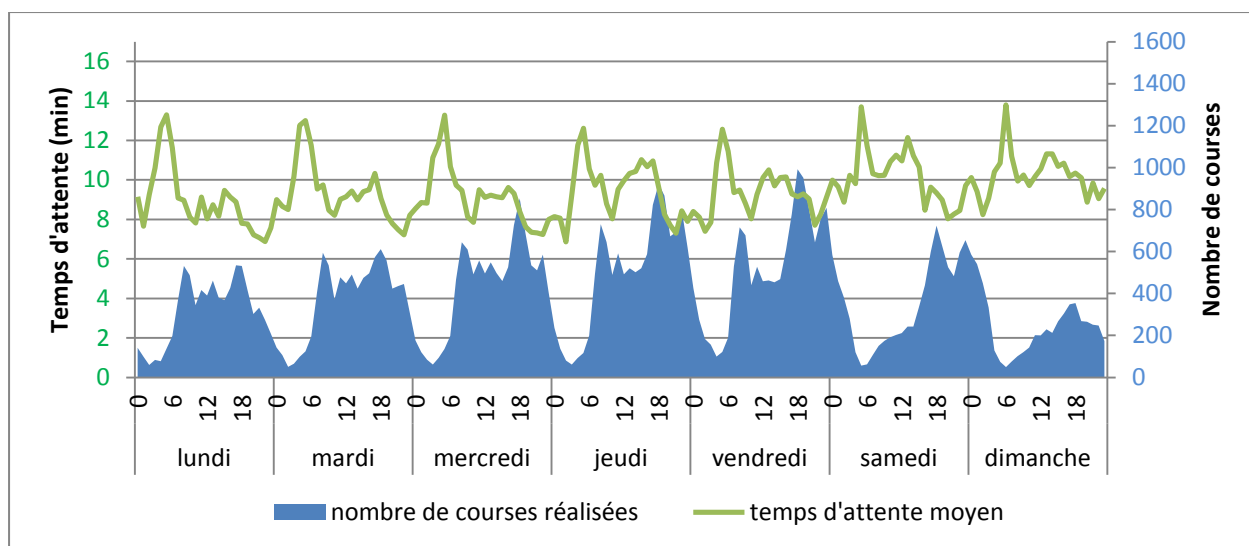


Figure 4-31 : Temps d'attente pour une semaine moyenne - Été

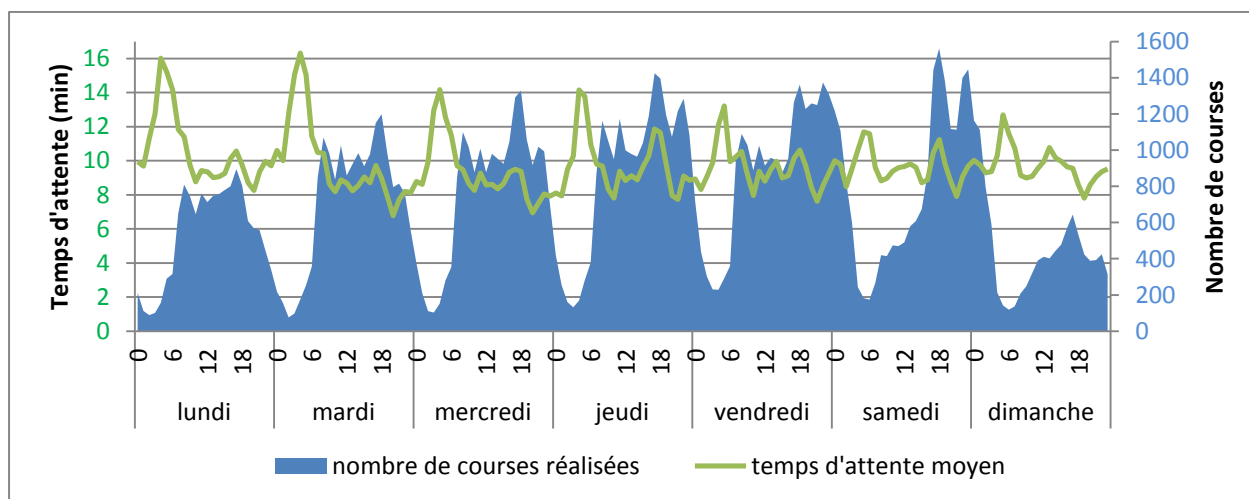


Figure 4-32 : Temps d'attente pour une semaine moyenne - Automne

Les figure 4-31 et figure 4-32 montrent l'évolution du temps d'attente par jour de semaine. Si le motif est semblable entre les deux saisons, le temps d'attente lui augmente en même temps que le nombre de courses entre l'été et l'automne. Il n'est pas en dessous des 6 min pour les 2 saisons et il atteint jusqu'à plus de 14 min en été et jusqu'à plus de 16 min en automne. Pour les deux saisons, un motif régulier se répète chaque jour de la semaine. Le temps d'attente atteint son maximum entre 3 et 6 h du matin même si le nombre de courses est faible. Puis il diminue et se stabilise durant l'après-midi entre 8 et 12 minutes selon les jours.

Il y a 45 128 usagers uniques sur la période juillet-décembre 2016 dont 20 620 en été et 36 825 en automne. Téo Taxi a servi en moyenne 690 usagers chaque jour d'été (coef. var. 33%) et 1347 usagers (coef. var. 32%) en automne. C'est-à-dire près de 1000 usagers chaque jour en moyenne pour les deux saisons. Tous les usagers ne font pas des courses tous les jours. Certains usagers peuvent commander des courses pour certaines occasions. Un usager actif est un usager ayant été servi par Téo Taxi. Ainsi, chaque jour, les usagers actifs font en moyenne 1,19 courses et ils annulent en moyenne 1,37 courses. Un usager attend en moyenne 9.3 min pour chacune de ses courses. La figure 4-33 montre la distribution cumulée du nombre de courses en fonction du pourcentage cumulé d'usagers ; chaque saison, 13% des usagers font à eux seuls 50% des courses et 50% des usagers font plus de 80% des courses.

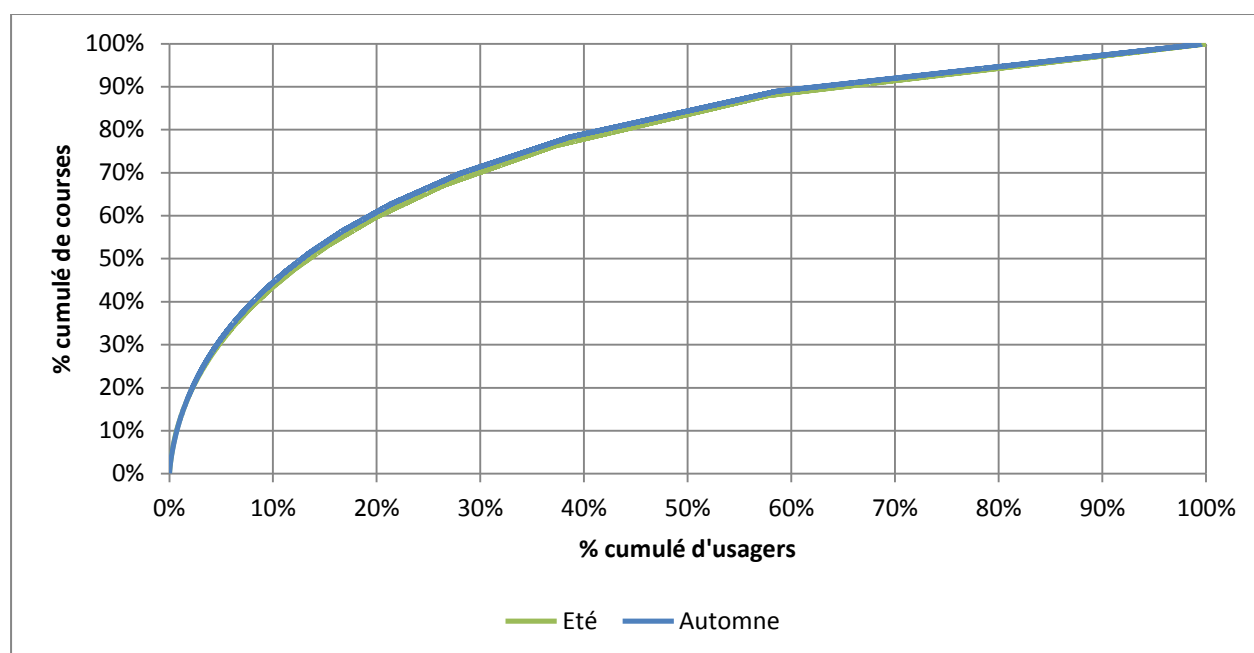


Figure 4-33 : Distribution cumulée du nombre de courses en fonction du pourcentage cumulé d'usagers

4.4 Conclusion

Pour résumer les faits saillants caractérisant l'offre et la demande, les chauffeurs rapportent près de 145\$ par jour. La majorité d'entre eux commencent à travailler le matin vers 6-7 h ou en début d'après-midi vers 13-15 h. À ce même moment, un grand nombre de chauffeurs finissent leurs quarts de travail, d'autres finissent plus tard vers 18-20 h. Tous les chauffeurs travaillant de jour sont en service près de 8 h par jour. Quant au taux d'utilisation des véhicules disponibles, il

tourne autour de 50% la journée et tend à diminuer vers 30% entre 2-5 h du matin. Le plus grand nombre de véhicules disponibles se trouve dans les mêmes secteurs de recensement que les sites de rechargement. En ce qui concerne la demande, durant la journée, elle est au plus fort le soir à partir de 17 h, elle se partage majoritairement entre le centre-ville et le Plateau Mont-Royal. La majorité des courses sont demandées par des usagers particuliers. Les courses aéroportuaires sont majoritaires entre 2 et 6 h du matin tout au long de la semaine. 50% des courses réalisées ont été demandées par seulement 13% des usagers de Téo Taxi.

CHAPITRE 5 IDENTIFICATION DE TYPOLOGIES DE JOURS DE DEMANDE

La construction de typologie - ou type - de jours de demande permet de caractériser la demande en taxi électrique en fonction des jours. Chaque type de jour représente une journée moyenne de déplacements en taxi électrique. Connaître les types de jours récurrents sur l'année, sur un territoire donné, permet à l'offre de s'organiser et satisfaire les demandes de courses émises chaque jour sur ce territoire. En effet si un chauffeur de taxi est en capacité de prévoir là où adviendra sa prochaine course, il pourra se positionner stratégiquement dans l'espace et dans le temps pour la satisfaire dans les plus brefs délais. À l'échelle d'une flotte de taxis, pouvoir prévoir la demande permet aux différents chauffeurs de coopérer pour maximiser la réalisation de leur service.

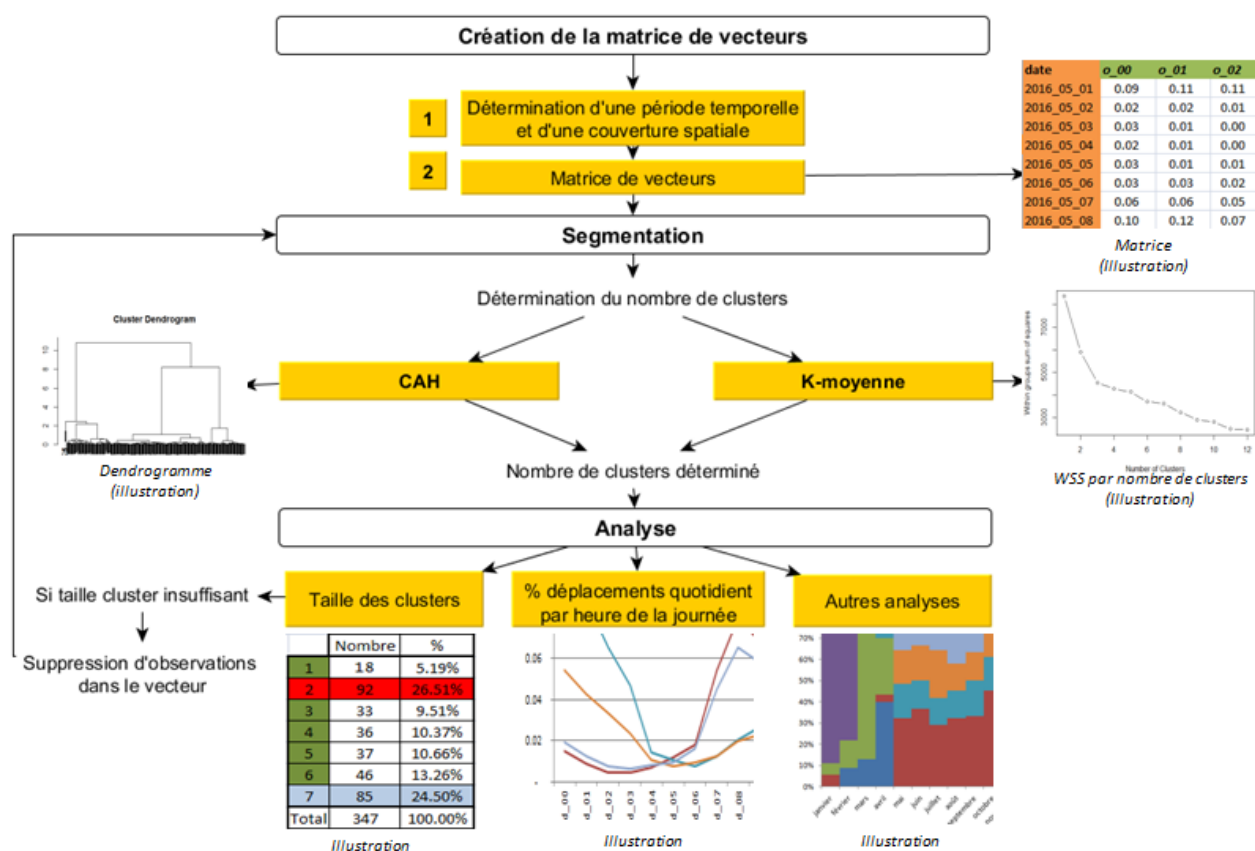


Figure 5-1 : Méthode de construction de typologie de jours de demande

Ce paragraphe décrit la méthodologie de construction de types de jours. Celle-ci peut se découper en 3 étapes : une première étape de création d'une matrice de vecteurs, une seconde étape de segmentation de cette matrice et une troisième étape d'analyse des clusters issus de la segmentation. La figure 5-1 décrit la méthodologie utilisée.

5.1 Construction d'une matrice de vecteurs

5.1.1 Types de vecteurs

Comme le montre la figure 5-2 plusieurs types de vecteurs ont été expérimentés : des *vecteurs jour*, des vecteurs *jour*zone* et des vecteurs *jour*zone²*. Une zone est un découpage administratif de la ville de Montréal comme le secteur de recensement ou le quartier d'habitation. Un ensemble de zones forme un territoire.

Types de matrices de vecteurs expérimentés			
Type de déplacements	Matrice de vecteurs jour	Matrice de vecteurs jour * zone	Matrice de vecteurs jour * zone ²
TERRITOIRE D'ETUDE	Couverture Téo au 1er janv	Couverture Téo au 1er janv	Couverture Téo au 1er janv
TEMPORALITE	jan/dec & avr/dec 2016	nov/dec 2016	nov/dec 2016
ZONAGE ADMINISTRATIF	-	SR & AR	SR & QR
DECOUPAGE HORAIRE	par heure	par heure	par heure & par période

Légende
SR : Secteurs de recensement
QR : Quartiers d'habitation
AR : Arrondissements

Figure 5-2 : Types de matrices réalisés

Un *vecteur jour* représente le pourcentage de déplacements quotidiens chaque heure de la journée, pour un jour et un territoire donnés. Il diffère du vecteur *jour*zone* qui lui représente, pour un jour et une zone administrative donnés, le pourcentage de déplacements quotidiens chaque heure de la journée. Quant au vecteur *jour*zone²*, il représente pour un jour et une paire

de zones administratives données, le pourcentage de déplacements quotidiens à chaque heure de la journée.

Plutôt qu'un découpage par heure de la journée, un découpage par période de la journée a aussi été essayé pour ce dernier vecteur. Les matrices de vecteurs seront plus amplement décrites dans les paragraphes qui suivent.

5.1.2 Définition du territoire et de la temporalité de l'étude

Parce que ces vecteurs dépendent du temps et de l'espace, il faut définir au préalable le cadre spatiotemporel de notre étude. Ce cadre délimitera le nombre de courses étudiées dans le temps et dans l'espace. Si la modélisation de la demande porte sur une partie de l'île de Montréal pour l'année 2016, alors l'étude porte sur α jours de services en taxis et notre territoire d'étude contient un nombre β de zones administratives où on eu lieu les origines et les destinations des courses. Construire une matrice de vecteurs revient à construire une matrice à $\alpha*\beta$ lignes où β diffère selon les cas.

5.1.3 Définition du type de déplacements

Les déplacements étudiés sont les déplacements originant et destinant. Ainsi à chaque heure (ou à chaque période) de la journée, un vecteur contiendra soit le pourcentage de déplacements originant soit celui des déplacements destinant, soit les deux.

Dans le cas où nous étudions séparément les déplacements originant et destinant, les matrices de vecteurs auront 24 colonnes, une pour chaque heure de la journée. Ou bien elles en auront 5, une par période de la journée (0-6 h, 6-9 h, 9-15 h, 15-18 h, 18-24 h). Au contraire, si ces deux types de déplacements sont étudiés ensemble elles en contiendront le double. La moitié pour les déplacements originant et l'autre moitié pour les déplacements destinant.

Les matrices construites sont résumées dans la figure 5-2 ci-dessus. La figure 5-3 ci-dessous montre un exemple de matrice de *vecteurs jour*SR* pour les déplacements originant. Ainsi à la date du "2016_11_01", il s'est réalisé dans le secteur de recensement 46202212, entre 4 h et 5 h du matin, 50% de ses déplacements originant.

date_sridu	o_00	o_01	o_02	o_03	o_04	o_05	o_06	o_07	o_08	o_09	o_10	o_11	o_12	o_13
2016_11_01_4620047	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2016_11_01_4620212	0	0	0	0	0.5	0	0	0.5	0	0	0	0	0	0
2016_11_01_4620351	0	0	0	0	0	0	0	0.17	0	0	0	0.17	0.17	0

Figure 5-3 : Exemple de matrice de vecteurs *jour*sr* pour les déplacements originant

5.2 Segmentation

5.2.1 La méthode des k-moyennes

Une fois la matrice de vecteurs est générée, il s'agit de segmenter les vecteurs en un certain nombre de clusters. Chaque vecteur sera alors associé à un cluster. La méthode de segmentation utilisée est la méthode des k-moyennes.

La méthode des k-moyennes est une méthode de segmentation permettant de partitionner un ensemble de données en k sous-ensembles (ou clusters) homogènes. Elle minimise la variance intracluster, nous permettant ainsi d'obtenir des clusters regroupant des données similaires. Appliquée à des données de jours de courses, cette méthode regroupe les jours de courses les plus similaires entre eux. Cependant avant de les regrouper, elle a besoin de savoir au préalable le nombre de clusters dans lesquels les données seront rassemblées.

Afin de pouvoir utiliser cette méthode, il est donc nécessaire de savoir déterminer ce nombre.

5.2.2 Dendrogramme et diagramme WCSS

Il est possible de déterminer le nombre de clusters à réaliser à l'aide d'un dendrogramme d'une part et d'autre part à l'aide d'un diagramme présentant la minimisation de la variance intracluster ou WCSS (pour Within Cluster Sum of Squares).

Le dendrogramme s'obtient à l'aide d'une autre méthode de segmentation : la méthode de classification ascendante hiérarchique (CAH). Cette méthode regroupe entre eux les éléments les plus similaires d'une base de données. Initialement, chaque élément constitue un cluster à lui tout seul. Le nombre de clusters est donc initialement égal au nombre d'éléments de la base de données. Puis en fonction d'une mesure de dissimilarité, les éléments sont regroupés entre eux en fonction de leur ressemblance diminuant ainsi le nombre de clusters. La méthode CAH fournit un

dendrogramme permettant alors de juger quel est le nombre minimal de clusters qui vient maximiser la variabilité intercluster et minimiser la variabilité intracluster.

Le diagramme WCSS minimisant la variance intraclusters en fonction du nombre de clusters est un second outil d'analyse. Il s'obtient par la réitération de la méthode des k-moyennes pour différents nombres de clusters.

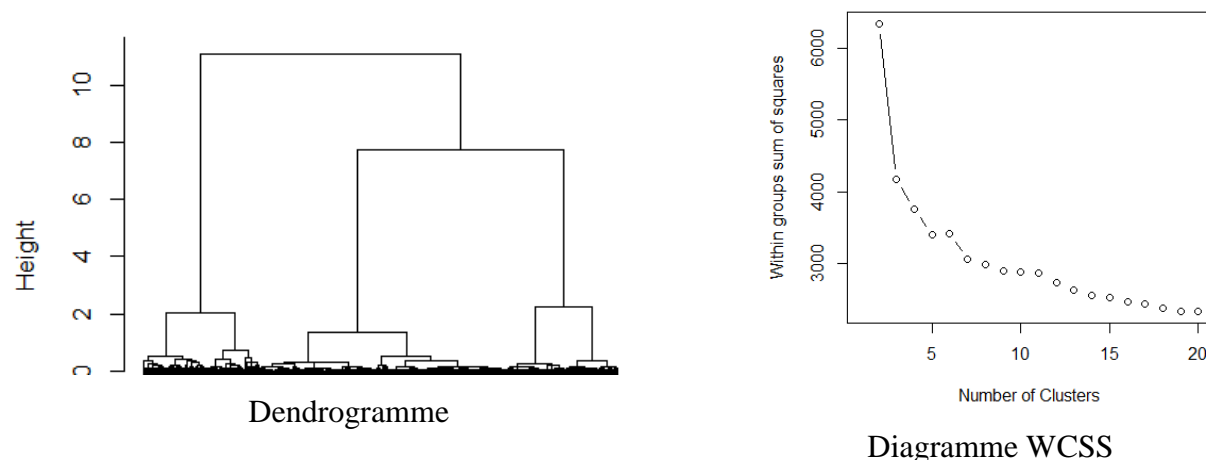


Figure 5-4 : Exemples de dendrogramme et de diagramme WCSS

Muni d'un dendrogramme par la méthode CAH et d'un diagramme WCSS, il est alors possible de déterminer le nombre de clusters minimal. La segmentation des vecteurs se fait selon ce nombre par la méthode des k-moyennes.

5.2.3 Analyses

L'étape d'analyse consiste à décrire les caractéristiques de chaque cluster. Plusieurs types de descriptions sont possibles : la description de la taille des échantillons, la description du vecteur moyen pour chaque cluster et la description de l'évolution de la prépondérance d'un cluster dans le temps et/ou dans l'espace.

5.3 Analyses des résultats

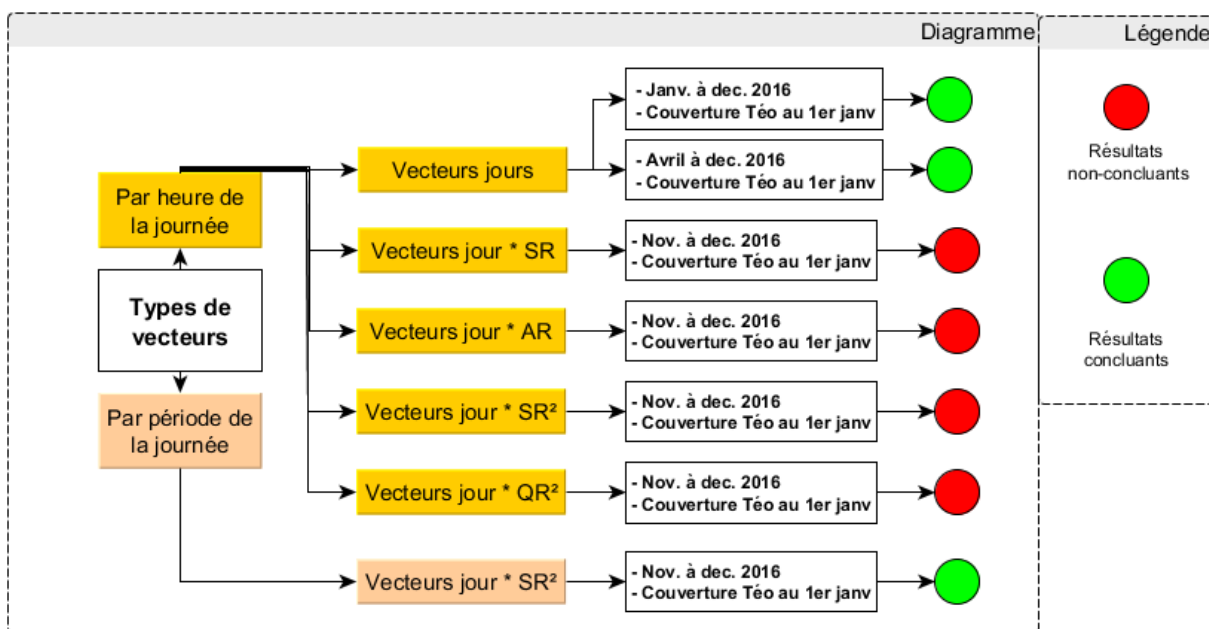


Figure 5-5 : Diagramme indiquant quels résultats ont été ou n'ont pas été concluants

La figure 5-5 indique quelles expérimentations ont été concluantes et lesquelles ne l'ont pas été. Comme nous le verrons par la suite, un résultat est concluant quand la variabilité intracluster de la segmentation donne un petit nombre de clusters (5-8 maximum) représentatif de jours-types de demande.

Pour tous les cas, le territoire d'étude est limité à la couverture spatiale du réseau de Téo Taxi en date du 1er janvier 2016 (cf. la figure 3-1). Afin d'avoir un nombre de vecteurs suffisants, il a été nécessaire pour la matrice de vecteurs jours de considérer toute l'année 2016. Pour les autres cas, seuls les mois de novembre et décembre 2016 ont été pris en compte. C'est aussi une période relativement stable, car comme le montre la figure 3-3 la couverture du réseau n'a pas évolué et le nombre de véhicules de la flotte n'a presque pas augmenté (+ 13%). Pour limiter le nombre de vecteurs vides, seuls les jours et les zones administratives où des courses ont eu lieu sont gardés.

5.3.1.1 Démarche générale de la recherche

Le but de cette recherche est d'identifier des types de jours de demande en prenant en compte l'évolution du nombre de courses dans le temps et l'espace. Ainsi il serait possible de prévoir la demande sur un territoire et à un temps donnés.

Tout d'abord nous pouvons ignorer la variable spatiale et chercher à identifier des types de jours en ne prenant en compte que l'évolution du nombre de courses dans le temps. Une matrice de vecteurs jours donne des résultats concluants. C'est-à-dire que par la méthode des k-moyennes nous obtenons un nombre raisonnable de types de jours de demande. Ce nombre si possible ne devrait pas dépasser 7 jours, soit un cluster par jour de semaine. Dans un premier temps, la période temporelle janvier-décembre 2016 avait été considérée. Cependant les premiers mois d'opération de Téo Taxi sont instables, car il y a eu une forte évolution de l'offre accompagnée d'une forte évolution de la demande. En effet la couverture du réseau a augmenté autour du 31 mars et la demande qui avait été près de 18 000 courses entre janvier et mars, est passée à près de 16 000 courses rien que pour le mois d'avril. C'est pourquoi l'expérimentation a été renouvelée sur la période d'avril-décembre 2016.

Afin de créer des types de jours de demande qui prennent en compte la composante spatiale du territoire, des vecteurs jour-SR ont été expérimentés dans un premier temps. Ainsi chaque vecteur caractériserait le pourcentage d'origines émises (ou le pourcentage de destinations attirées) par un secteur de recensement pour chaque heure de la journée. Cependant du fait du nombre trop important de clusters à fabriquer, les résultats n'ont pas été concluants. Il a alors été expérimenté les vecteurs jour-sr². Chaque vecteur dès lors caractérisait le pourcentage de déplacements quotidiens pour chaque paire OD et pour chaque jour de la période. Les résultats n'ayant pas été concluants de nouveau, un découpage administratif plus grand a été essayé : les quartiers d'habitation de Montréal (qr). Une matrice de vecteurs jour*qr² a été segmentée cependant les résultats n'ont toujours pas été concluants. Le nombre de clusters à fabriquer était trop grand.

L'idée suivante fut de regrouper les heures de la journée afin de créer des périodes. Ainsi 5 périodes ont été construites : de 0-6 h, 6-9 h, 9-15 h, 15-18 h, 18-24 h. La nouvelle matrice était donc composée de vecteurs jour*sr² par groupe d'heures.

La figure 5-6 ci-dessous montre un exemple de matrice de *vecteurs jour*sr² par groupe d'heures*. Ainsi à la date du "2016_11_01", 100% des déplacements quotidiens depuis le secteur de

recensement 4620041 en destination du secteur de recensement 4620062 ont été réalisés entre 6 et 9 h du matin.

date_pairOD	X0_6	X6_9	X9_15	X15_18	X18_24
2016_11_01_4620041_4620062	0	1	0	0	0
2016_11_01_4620041_4620163	0	0	0	0	1
2016_11_01_4620042_4620051	0	0	1	0	0

Figure 5-6 : Exemple de matrice de vecteurs $jour*sr^2$ par groupe d'heures

5.3.1.2 Résultats

Matrice de vecteurs jours

Pour la période d'avril-décembre 2016, la matrice de vecteurs jours contient 275 vecteurs. La segmentation les a regroupés en 4 clusters.

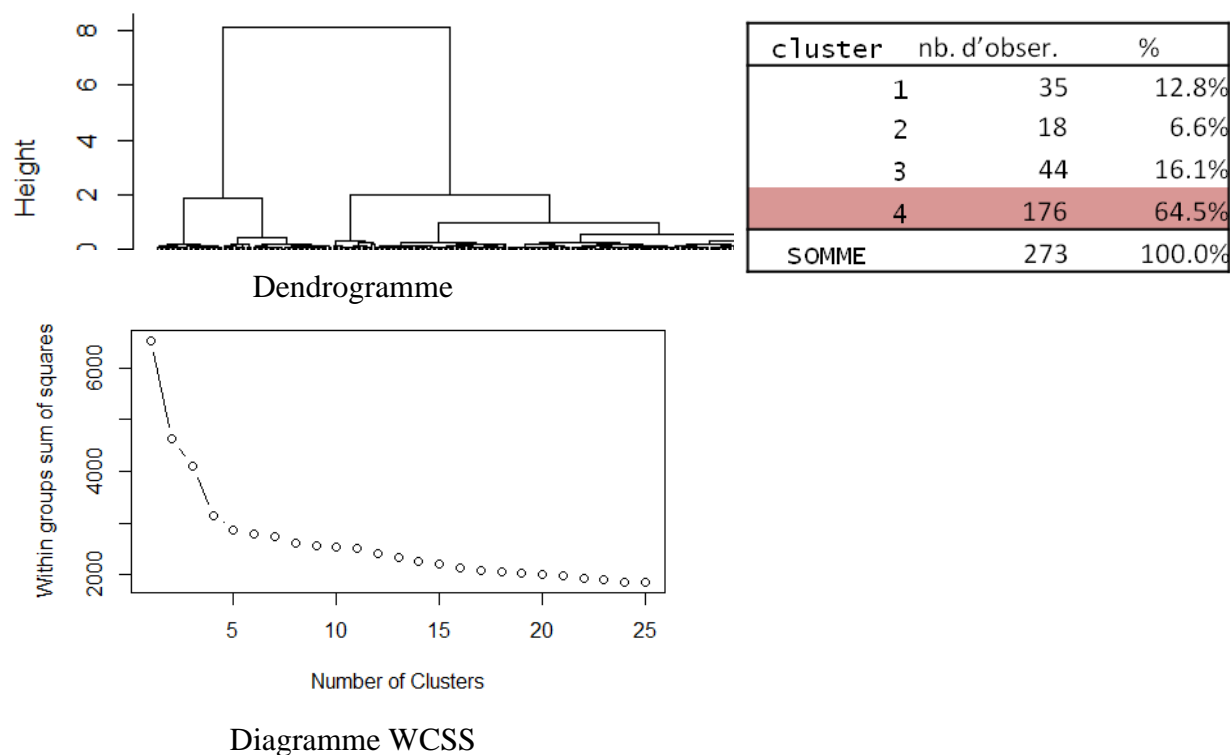


Figure 5-7 : Évaluation du nombre de clusters (matrice de vecteurs *jours*)

Pour la période d'avril-décembre 2016, la matrice de vecteurs jours contient 275 vecteurs. Le dendrogramme et le diagramme WCSS déterminent le nombre de clusters dans lesquels ils seront

regroupés. Ici 4 clusters sont suffisants. Le groupe numéro 2 contient le moins de vecteurs (7%) et le groupe numéro 4 en contient le plus (64%).

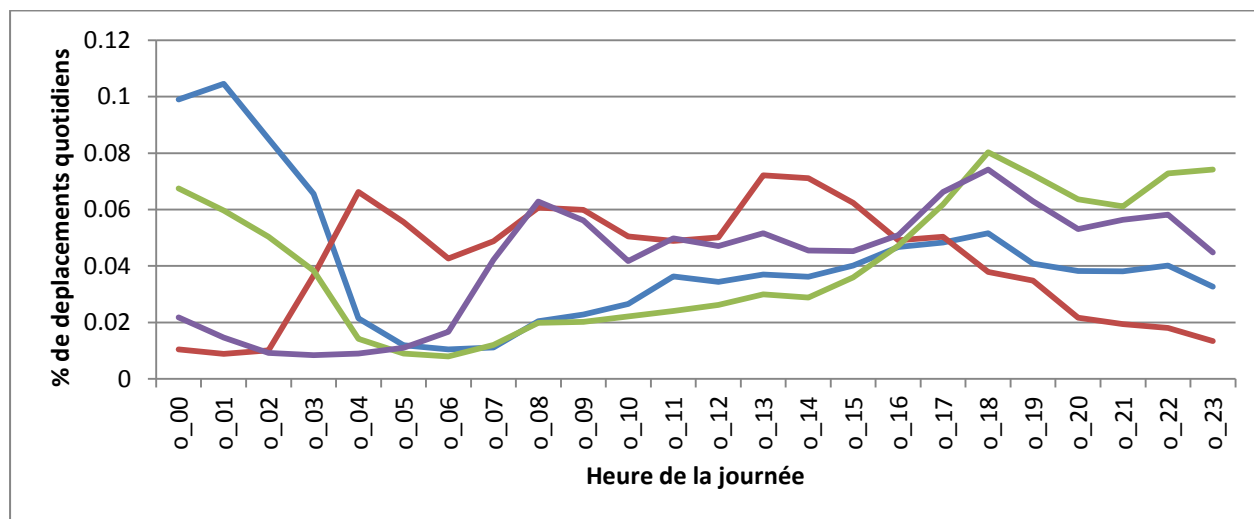


Figure 5-8 : Pourcentage de déplacements quotidiens par heure de la journée, par type de jour

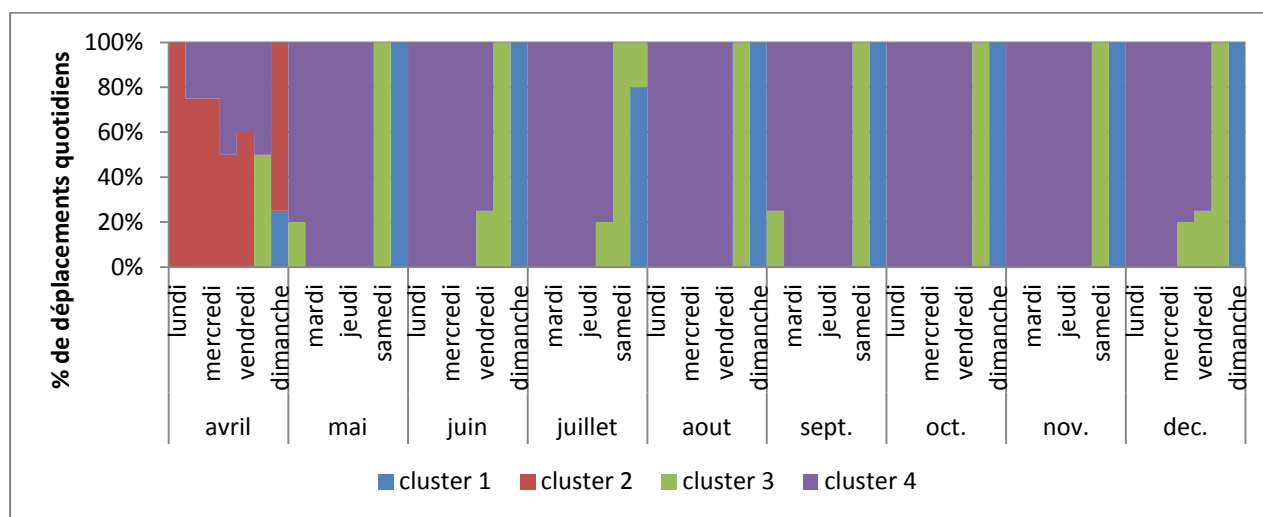


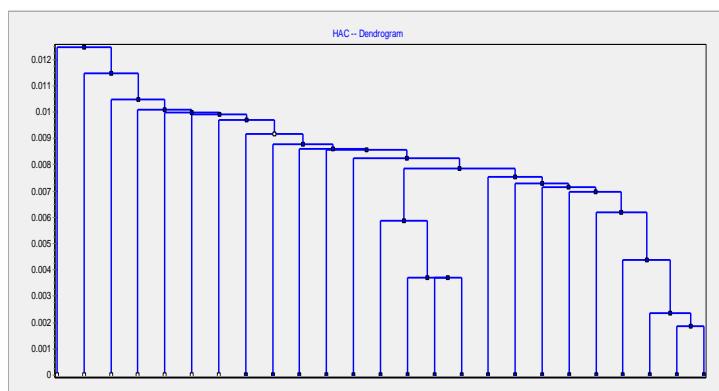
Figure 5-9 : Types de jours dominants par jour de l'année 2016 (vecteur *jours*)

La figure 5-9 nous montre que 3 types de jours de demande sont prépondérants d'avril à décembre 2016 : les clusters violet, vert et bleu qui représentent respectivement le cluster des jours ouvrables, le cluster des samedis et le cluster des dimanches. Le vendredi, le cluster vert est aussi présent à 25% ou moins. Cela signifie que, durant le mois, la journée de demande de certains des vendredis ressemble à la journée de demande des samedis. La figure 5-8 donne les vecteurs moyens issus de la segmentation. Le vecteur des jours ouvrables (violet) montre que le pourcentage de déplacements quotidiens connaît 2 pics : l'un à 8 h du matin et l'autre à 18 h. Le

premier pic correspond à 6% des déplacements quotidiens, le second à près de 8%. Ces pics correspondent aux heures de pointe typiques des journées ouvrables. Ensuite le cluster des samedis (vert) connaît 2 pics. L'un très tôt le matin, à 0 h et le second plus diffus est à 18 h. Le premier pic correspond à plus de 6% des déplacements quotidiens, le second à 8%. Les usagers du taxi le samedi sont probablement des gens qui viennent de finir leur soirée du vendredi soir ou qui vont commencer leur soirée du samedi soir. Le cluster du dimanche (bleu) ressemble au cluster du samedi, probablement parce que les usagers du taxi pour ces deux jours se déplacent avec des motifs similaires. Le cluster du dimanche connaît un plus fort pic que le samedi matin à 1h avec près de 10% des déplacements quotidiens effectués. Ce pic chute à 4 h du matin à moins de 2% et la courbe se rehausse doucement avec un nouveau pic moins élevé que le samedi à 18h avec environ 5% des déplacements quotidiens réalisés.

*Vecteur jour*zone*

Après avoir constaté qu'il était possible de générer des jours-types de demande avec la donnée temporelle seulement, il s'agit d'essayer d'ajouter la donnée spatiale.



Dendrogramme

cluster	nb. obser.	%
1	7859	40%
2	1055	5%
3	1066	5%
4	5880	30%
5	1174	6%
6	712	4%
7	511	3%
8	1603	8%
SOMME	19860	100%

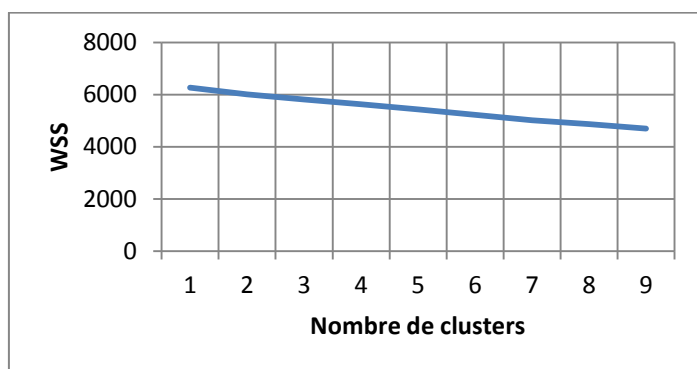


Diagramme WCSS

Figure 5-10 : Évaluation du nombre de clusters (matrice de vecteurs *jour*SR d'origine*)

Les nouveaux vecteurs étudiés sont les vecteurs *jour*zone*. Précisément, ce sont les vecteurs *jour*SR d'origines*. La figure 5-3 donne une illustration des vecteurs utilisés. Les diagrammes figure 5-10 ne permettent pas de déterminer clairement le nombre de clusters à réaliser. 4 ou 8 clusters pouvaient être réalisés d'après le dendrogramme. Afin de diminuer l'erreur intracluster, ce sont 8 clusters qui ont été faits. La figure 5-11 montre les vecteurs moyens associés à ces clusters. La trop grande variabilité intracluster engendre des clusters avec peu d'observations. Les clusters 1 et 4 présentés à la figure 5-12 représentent à eux deux 70% des observations.

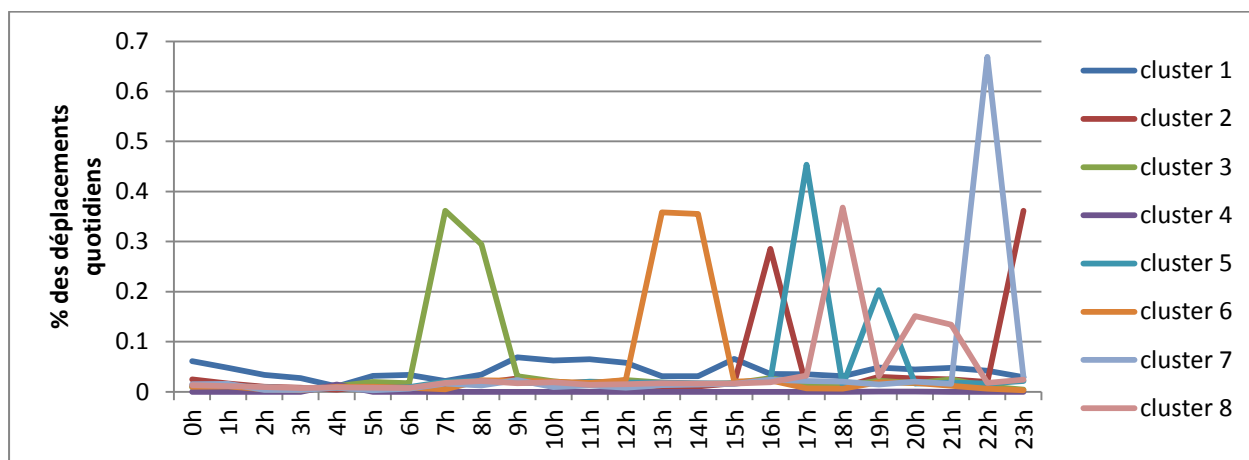


Figure 5-11 : Pourcentage de déplacements quotidiens par heure de la journée, par type de jour

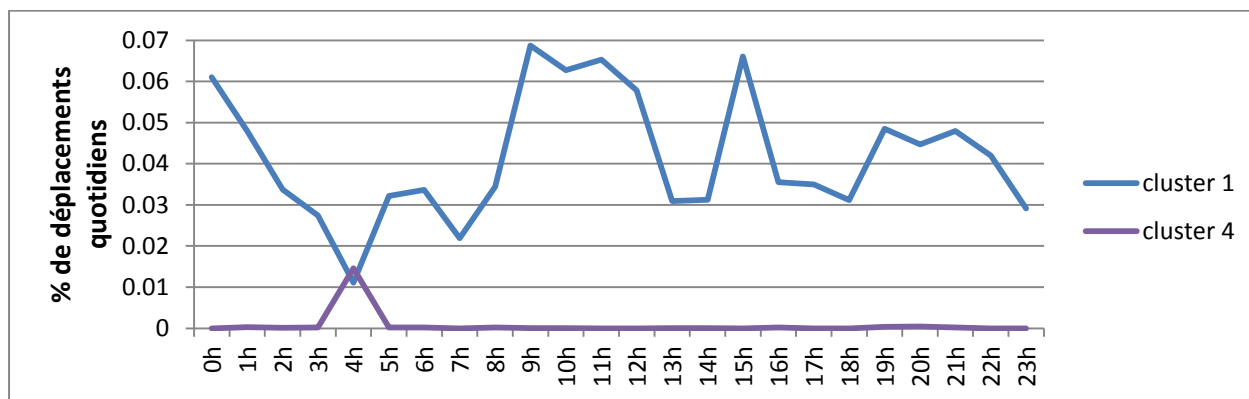


Figure 5-12 : Pourcentage de déplacements quotidiens par heure de la journée, pour les clusters 1 et 4

Les résultats de cet essai ne sont pas concluants du à la trop grande variabilité intracluster. Des résultats aussi peu concluants sont obtenus si le territoire est découpé en arrondissements plutôt qu'en secteurs de recensement.

*Vecteur jour*zone²*

L'idée nouvelle était alors non pas de faire des vecteurs *jour*zone*, mais des vecteurs *jour*zone²*. C'est-à-dire des vecteurs où le pourcentage quotidien par heure de la journée se ferait par jour et par paire OD. Cette approche n'a pas marché. Une dizaine de clusters a été réalisée par segmentation, mais la variabilité intracluster reste trop grande. Cependant l'approche suivante a donné de meilleurs résultats : étudier les déplacements quotidiens par *groupe d'heures* et par paire OD. Ce sont donc ces résultats que nous présentons.

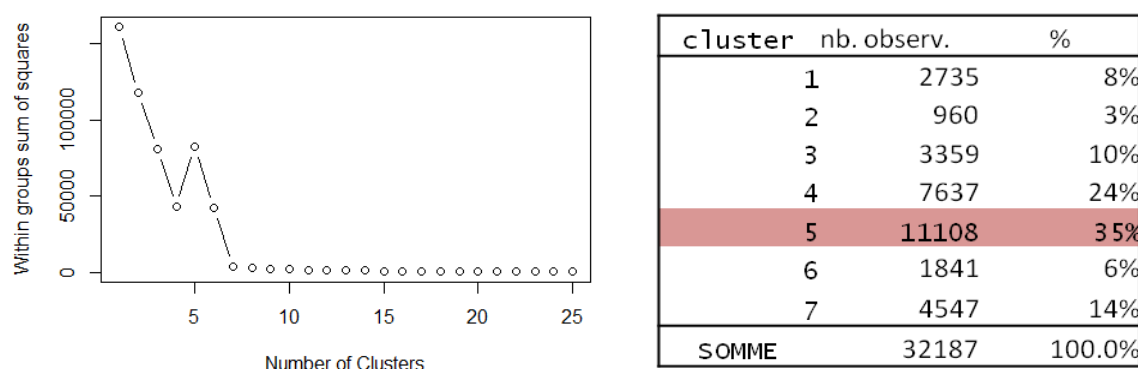


Figure 5-13 : Évaluation du nombre de clusters (*jour*paire OD* par groupe d'heures)

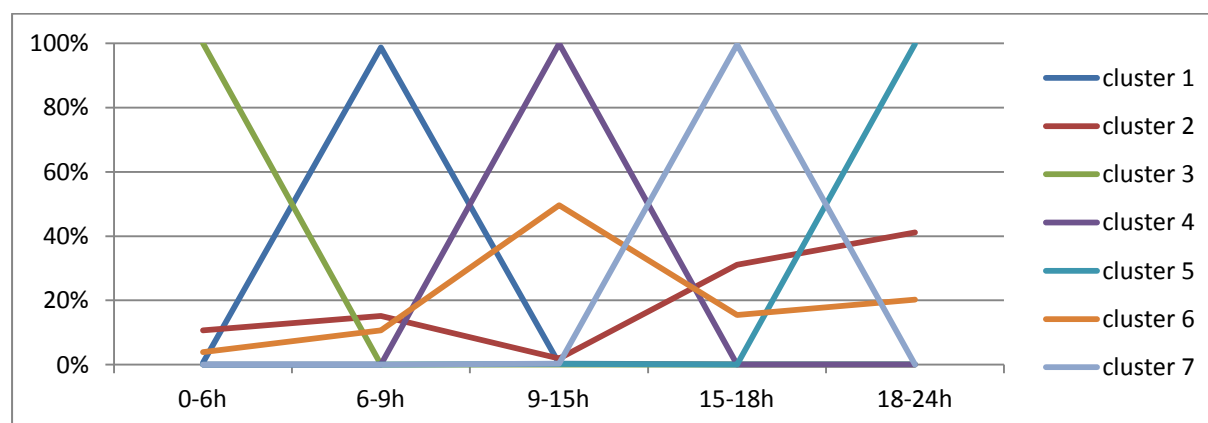


Figure 5-14 : % de déplacements quotidiens par heure de la journée, par type de jour (*jours*sr²* par groupe d'heures)

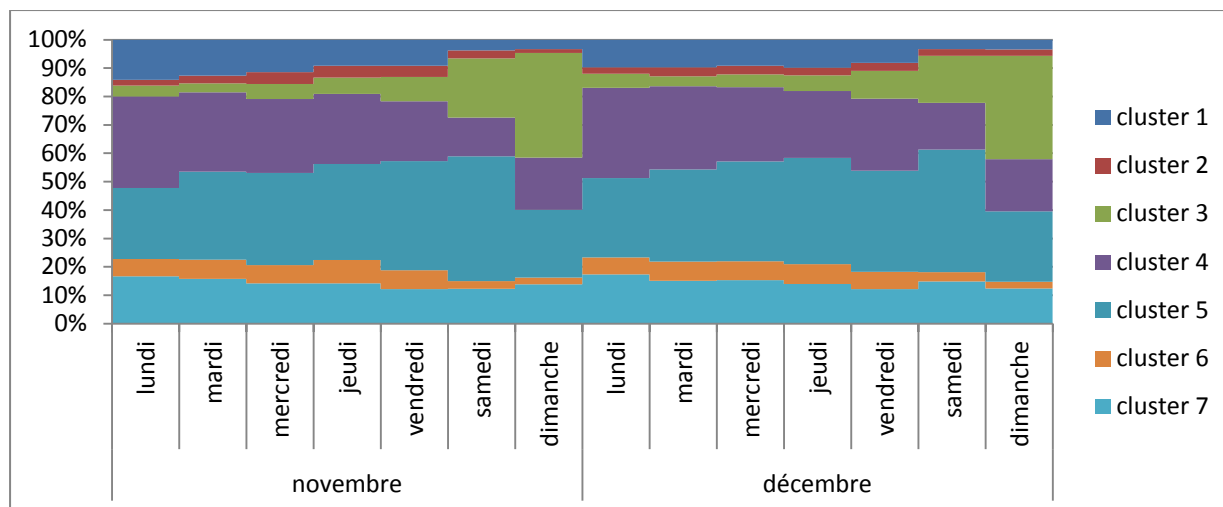


Figure 5-15 : Types de jour dominants par jour de l'année 2016 (jours*sr² par groupe d'heures)

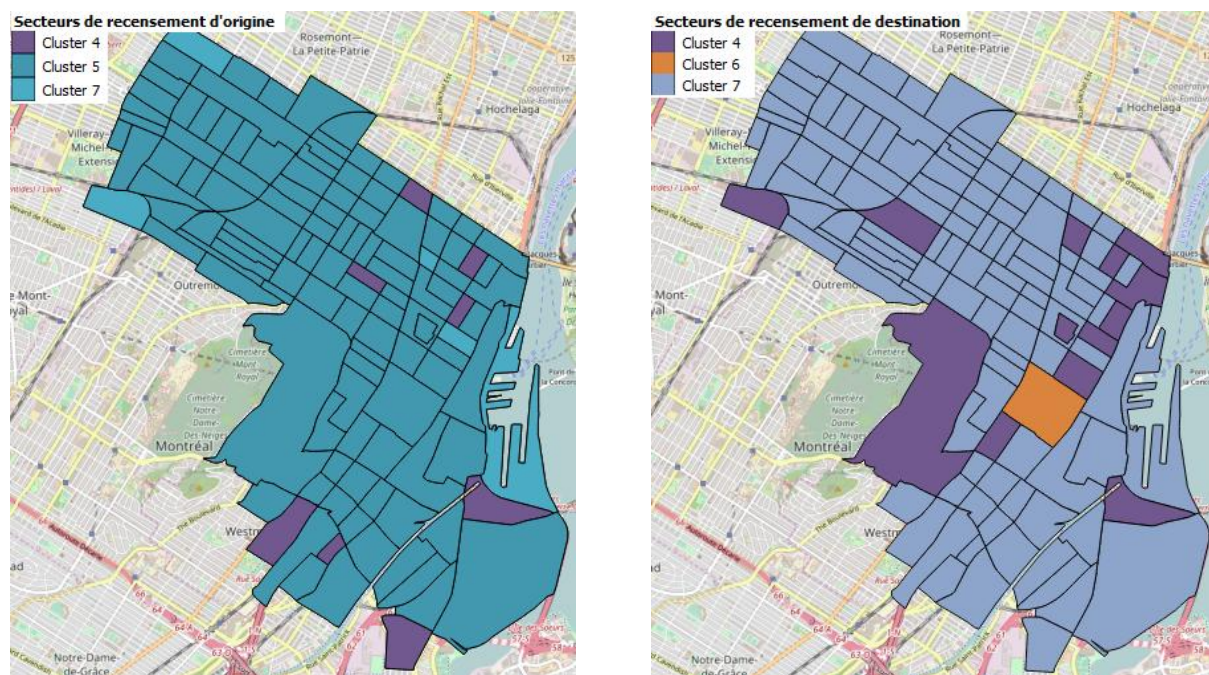


Figure 5-16 : Clusters dominants pour les secteurs d'origines et de destinations

D'après la figure 5-13, 7 clusters suffisent. La figure 5-15 nous montre que contrairement aux vecteurs *jour*, les clusters ne correspondent pas à des jours-types mais à des périodes horaires types qui sont plus ou moins prépondérantes en fonction du jour de la semaine.

Ainsi le cluster violet qui correspond à 100% de déplacements quotidiens durant la période horaire 9-15 h est prépondérant le lundi. Le dimanche, c'est le cluster vert qui correspond à 100% des déplacements quotidiens de la journée durant la période horaire 0-6 h. Les clusters rouge et

orange qui semblent être respectivement le pourcentage de déplacements durant les heures de pointe et durant les heures creuses ne sont pas prépondérants tout au long de la semaine.

La figure 5-16 montre les clusters dominants pour les secteurs d'origines et de destinations. Ainsi le cluster 5, qui correspond aux déplacements du soir entre 15-18 h, domine pour les secteurs d'origine. C'est-à-dire que la majorité des déplacements faits dans ces secteurs dominés par le cluster 5 se font entre 15-18 h. De même les secteurs de destinations sont dominés par le cluster 7 qui correspond aux déplacements faits entre 18-24 h. La majorité des déplacements destinant de ces secteurs se font donc durant cette période. Ces deux cartes illustrent le phénomène des usagers qui sortent le soir entre 15-18 h et rentrent chez eux entre 18-24 h. Cet outil cartographique comporte néanmoins une limite : pour un même secteur de recensement, certains clusters se concurrencent, mais il n'est affiché que celui qui domine de peu.

Conclusion

La segmentation des matrices de vecteurs donne de clairs résultats lorsque les vecteurs de déplacements quotidiens dépendent uniquement de la donnée temporelle. Lorsque la donnée spatiale est ajoutée, la recherche s'avère difficile. Les vecteurs jour*zone qui font dépendre les déplacements seulement d'une extrémité de leurs extrémités ne permettent pas d'identifier de jours-types de demande à cause d'une trop grande variabilité intracluster durant la segmentation. Pour cette même raison, aucun résultat n'est obtenu lorsque qu'on utilise des vecteurs jours*zone² par heure de la journée. Néanmoins lorsque les heures sont regroupées par période, la variabilité intracluster diminue et des résultats commencent à être obtenus. Aucun jour-type de demande ne ressort. Les clusters ne représentent seulement que des périodes de demande que l'on retrouve plus ou moins dominante au cours des jours de la semaine.

CHAPITRE 6 ALGORITHME DE GÉNÉRATION DE L'OFFRE DE TAXIS ÉLECTRIQUES

Le chapitre 4 a exposé une certaine manière d'étudier les comportements de mobilité des usagers se déplaçant par taxi électrique, ainsi que les caractéristiques propres à une telle offre de transport. Une autre façon d'étudier l'offre et la demande d'un système de transport par taxi électrique est la modélisation. Modéliser les différents objets (chauffeurs, véhicules, usagers, etc.) et leurs interactions peut permettre de prévoir le comportement d'un tel système selon différents scénarios. C'est de cette manière que Maciejewski (2014) a étudié, à l'aide d'un logiciel orienté-agent, l'impact de l'augmentation de la demande ou encore d'une diminution du nombre de bornes de chargement sur le réseau de taxis électriques dans une ville fictive.

Ce chapitre a pour but de préparer la modélisation d'un système de transport par taxi électrique. Dans un premier temps, nous présentons les différents paramètres qualitatifs et quantitatifs ayant un impact sur une éventuelle modélisation. Dans un second temps, un algorithme de génération de l'offre est proposé et expliqué. À partir d'une demande donnée, cet algorithme génère un nombre de véhicules et un nombre de bornes adéquates pour harmoniser l'offre avec la demande.

6.1 Paramètres quantitatifs et qualitatifs

6.1.1 Modélisation du territoire d'étude

Le territoire urbain d'étude est l'île de Montréal. Nous pourrions découper le territoire selon un quadrillage où chaque case serait une zone du territoire avec ses spécificités. Ces cases pourraient fusionner avec leurs voisines en fonction de leur ressemblance, donnant alors des zones plus ou moins grandes. Les spécificités de chaque zone caractériseraient l'état de la demande et ce qui l'influence, ainsi que l'état de l'offre et ce qui l'influence comme l'illustre la figure 6-1.

Spécificités des zones	
Demande	Nombre de résidences, nombre d'emplois, etc.
Offre	Degré des pentes, températures, etc.

Figure 6-1 : Exemples de spécificités caractérisant chaque zone du territoire

6.1.2 Les usagers

La demande est l'ensemble des courses formulées par les usagers. Ce sont donc ces derniers qui décident d'où et quand auront lieu les courses. Les usagers peuvent réaliser leurs requêtes d'au moins 2 façons : par réservation ou par hélage. Les requêtes par réservation peuvent prendre plusieurs formes : certaines peuvent être faites par téléphone, d'autres par application mobile ou encore directement par portail internet. La demande se caractérise par, a minima, le nombre de courses qui émanent des usagers, où et quand elles émanent, leurs origines et destinations.

6.1.3 La flotte de taxis

Différents paramètres définissent une flotte de taxis. Le nombre de taxis nécessaire pour répondre à la demande doit être déterminé, mais aussi quel(s) type(s) de taxis composera la flotte. Cette dernière pourra d'ailleurs ne pas être entièrement électrifiée. Elle serait alors en partie composée de taxis par exemple hybrides. La batterie de chaque taxi électrique est au cœur du problème. La dynamique du réseau de taxis dépendra du *niveau de charge* de la batterie de chaque taxi et de son autonomie. Il émerge donc deux contraintes : le chauffeur de taxi doit régulièrement retourner à une borne de recharge, et doit effectuer le remplacement de la batterie lorsqu'elle n'est plus suffisamment autonome. Par ailleurs l'hiver canadien affectant la batterie, il sera important de prendre en compte l'évolution de l'autonomie d'une batterie en fonction du climat montréalais.

6.1.4 Définition des seuils d'autonomie de la batterie

Soit **B** le niveau de charge de la batterie quand elle est pleinement chargée. Nous pouvons scinder **B** en plusieurs seuils.

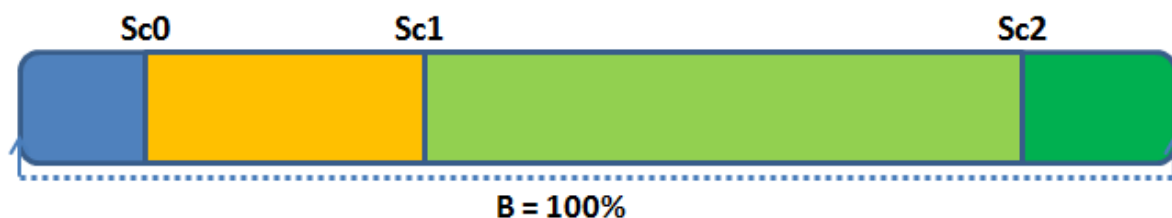


Figure 6-2 : Description des seuils de la batterie électrique

Nous posons **Sc1** comme étant le seuil à partir duquel un taxi doit aller se recharger. Cependant le taxi peut encore prendre des usagers qui sont sur le chemin qui le sépare de la borne de recharge. Si par exemple, l'origine et la destination de la course d'un usager se trouvent sur la trajectoire qui le sépare de sa prochaine borne de recharge, alors il peut réaliser cette course. Nous pouvons nommer une telle course, une *course transparente*. Au contraire si soit l'origine, soit la destination de la course est en dehors de cette trajectoire, alors le taxi devra refuser cette course.

Nous pouvons scinder la batterie en plusieurs seuils. Avant **Sc0** : aucune course n'est possible, le taxi doit impérativement aller se recharger à la borne la plus proche. Entre **Sc0-Sc1** : des courses transparentes sont possibles, c'est-à-dire des courses qui sont sur le chemin qui le sépare de la borne de recharge. Entre **Sc1-Sc2** : de courtes et longues courses sont possibles, et il est possible de ne pas recharger au-delà du seuil **Sc2**. Entre **Sc2-Sc3** : de très longues courses sont possibles, la batterie est pleine. Chaque seuil correspond à un pourcentage de l'alimentation de la batterie, pourcentage qui est à définir.

6.1.5 Les bornes de recharge

Le problème de l'établissement d'un réseau de taxis électriques a des similitudes avec le problème de l'établissement d'un réseau de taxis traditionnels (à essence). La différence principale réside dans le fait que les taxis traditionnels ont pour postes d'alimentation les stations d'essence dont les infrastructures sont déjà existantes et intégrées à l'espace urbain. Tandis que les taxis électriques ont besoin de stations de recharge reliées au réseau électrique urbain dont les infrastructures ne couvrent pas à ce jour l'ensemble du territoire de l'île de Montréal. Ainsi dans le problème de l'établissement d'un réseau de taxis électriques, il ne s'agit pas seulement de trouver l'emplacement des postes d'attentes des usagers, il faut encore réfléchir à l'emplacement adéquat des bornes de recharge à venir. De telles bornes pourront par ailleurs servir à de futurs véhicules électriques privés particuliers.

Le nombre de bornes qui doivent être implantées doit être déterminé. Le type de chaque borne doit être aussi défini. Le type d'une borne est caractérisé par sa vitesse de chargement, mais aussi par la technologie utilisée : est-ce une borne de chargement par branchement, ou une borne de rechargement par induction. Il reste à déterminer les critères qui détermineront qu'une borne sera d'un certain type plutôt que d'un autre. A priori, nous pouvons poser que la fréquence d'utilisation

d'une borne déterminera son type. Cette fréquence d'utilisation sera probablement impactée par la densité de la demande de la zone dans laquelle elle se trouve. L'emplacement des bornes de recharge est donc aussi à déterminer. Cet emplacement sera conditionné notamment par la possibilité d'acheter ou de louer le terrain sur lequel la borne sera installée, et par la possibilité de la connecter au réseau électrique urbain.

6.1.6 Les postes d'attente

Le nombre de postes d'attente est à déterminer, ainsi que leur emplacement. L'emplacement d'un poste d'attente sera notamment déterminé par la localisation de la demande en taxi sur le territoire, ainsi que la rentabilité d'un service de taxis dans la zone couverte par ce poste. Les bornes de recharge pourront être distinctes et distantes des sites d'emplacement des postes d'attente.

6.2 Algorithme de génération de l'offre

Dans ce paragraphe il est proposé une méthodologie de génération de l'offre en taxi électrique. Ce processus est inspiré de la méthode de Kara Kockleman, (2016). Plus précisément, pour une demande donnée, un processus permettant de décider du nombre de taxis et de bornes de rechargement est décrit. La figure 6-3 ci-dessous résume le processus. Le processus décrit par la cette figure est en 2 phases. La première sert à produire le nombre de taxis électriques et de bornes de recharge pour une demande donnée. La seconde sert à expliquer le fonctionnement d'un réseau de taxis électriques avec un nombre véhicules et de bornes données.

Plaçons-nous durant une journée type. Cette journée se compose de 2 périodes d'heures de pointe, celle du matin et celle du soir. Ainsi que 3 périodes d'heures creuses. Supposons qu'une des deux périodes d'heures de pointe connaisse une plus forte demande que l'autre. Alors trouver le nombre théorique de taxis électriques, ainsi que le nombre de bornes électriques qui les assisteront, revient à trouver leur nombre théorique requis durant cette période de pointe.

Supposons-nous au début de cette période de pointe, à un temps initial. À ce temps initial, les taxis électriques sont tous à une position géographique initiale (garage, poste d'attente, position stratégique, etc.).

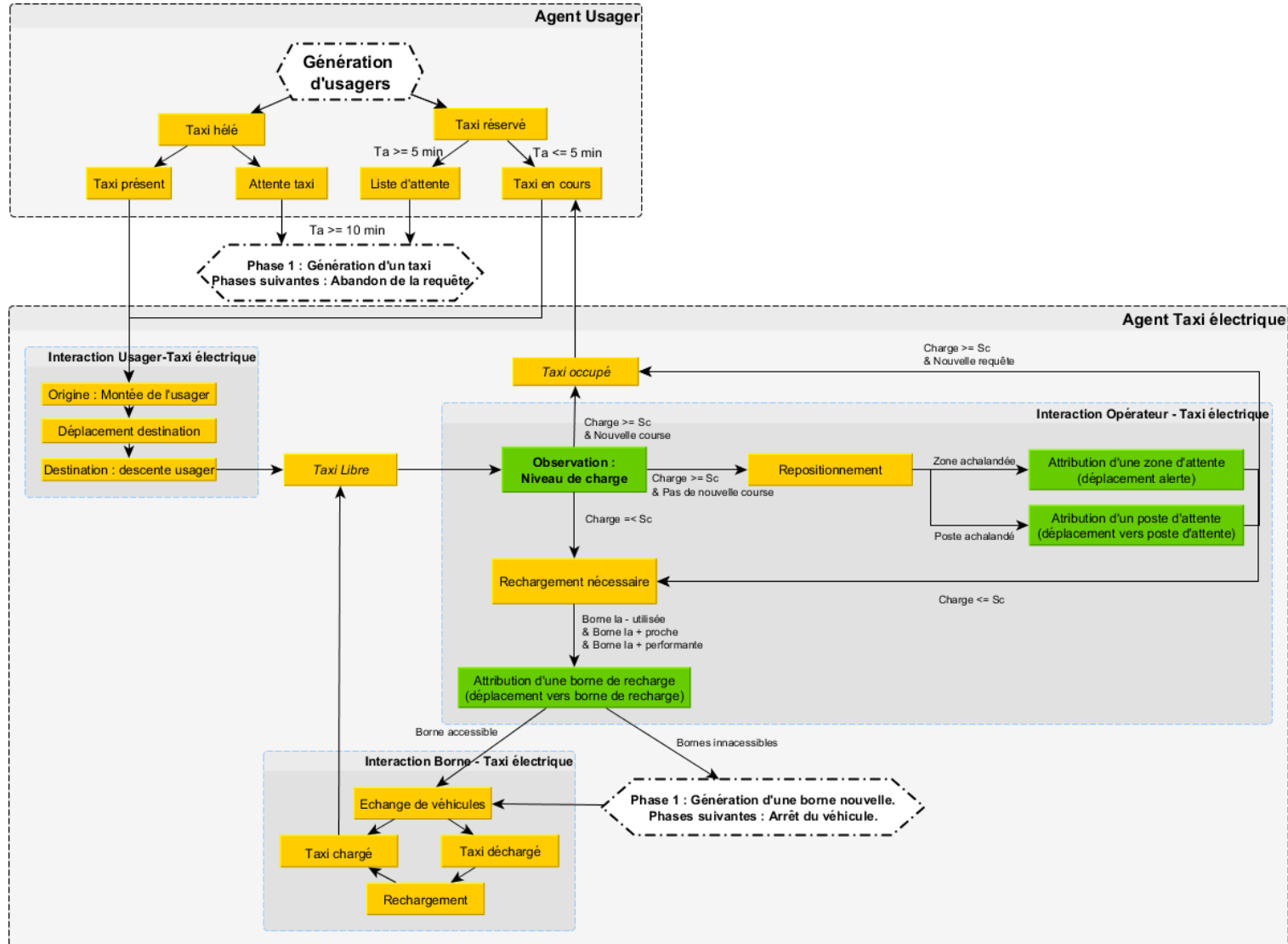


Figure 6-3 : Description orientée agent d'un réseau de taxis électriques

La figure 6-3 fait une description orientée agent d'un réseau de taxis électriques. Les sections ci-après présentent les différents points de vue des agents représentés sur la figure.

6.2.1 Point de vue de l'agent usager

À un instant de la période de pointe et à un lieu donnés du territoire, un usager fait une demande de course en taxi. Il a deux manières de la présenter :

- *L'usager fait une réservation à l'aide d'une requête.* Alors si un taxi peut satisfaire la course dans les 5 min qui suivent, l'usager attendra son taxi durant un temps d'attente Ta inférieur à 5 min avant de monter à bord du véhicule. Sinon, si aucun taxi ne peut satisfaire cette course dans les 5 min qui suivent, le temps d'attente Ta de l'usager sera donc supérieur à 5 min et sa demande tombera dans la liste d'attente. Si toujours aucun taxi ne peut réaliser la course en liste d'attente dans les 10 min suivant la requête, il y a *génération d'un taxi électrique*.
- *L'usager hèle un taxi.* Dans ce cas, le taxi peut accepter ou pas de le prendre selon sa disponibilité.

6.2.2 Point de vue de l'agent taxi électrique

Le conducteur de taxis électriques reçoit une requête. Son véhicule passe en mode *occupé*. Il se déplace jusqu'à l'usager pendant son temps d'attente Ta . Une fois que l'usager est monté à bord du véhicule, le taxi le conduit jusqu'à sa destination où il descendra.

Le taxi passe alors en mode *libre*. Il vérifie ensuite si son niveau de charge est au-dessus ou au-dessous d'un certain seuil critique **Sc1**. Trois options s'offrent à lui :

- Son niveau de charge est au-dessus du seuil critique **Sc1** et il a une nouvelle course à réaliser. Alors il s'y rend. Son statut passe de nouveau en mode *occupé*.
- Son niveau de charge est au-dessus du seuil critique **Sc1** et il n'a pas de nouvelle course à réaliser. Alors deux nouvelles options s'offrent à lui :

- Une zone d'achalandage lui est attribuée. Le taxi va alors se déplacer de manière alerte et attendre qu'un usager le hèle ou de recevoir une commande de course. Si aucun ne le hèle et que son niveau de batterie retombe au seuil **Sc1**, il retourne se recharger.
- Un poste d'achalandage lui est attribué. Le taxi va alors se déplacer jusqu'au poste et attendre un usager si ce dernier n'est pas déjà là.
- Son niveau de charge est au-dessous du seuil critique **Sc1**, alors il se dirige vers la borne de recharge qu'on lui a attribuée et grâce à laquelle il se rechargera. Deux situations adviennent :
 - Si aucune borne n'est à proximité, une borne est générée et positionnée à l'endroit le plus approprié.
 - Il peut réaliser une *course transparente* si son niveau de charge est au-dessus du seuil Sc0.

6.2.3 Facteurs de décisions

Dans l'algorithme, le chauffeur de taxi doit faire deux choix : un choix de repositionnement (à un poste ou une zone d'attente) et un choix de borne de recharge.

6.2.3.1 Choix du repositionnement

Un taxi libre n'ayant aucune course à réaliser a le choix de se repositionner à un poste d'attente ou à une zone d'attente achalandée.

Les emplacements des postes d'attente sont prédéfinis. Les zones d'attente correspondent à des zones où l'achalandage est potentiellement fort et réparti sur toute une zone plutôt que concentré autour d'un lieu. Une zone d'attente contient ou non des postes d'attente. Nous pouvons attribuer aux postes et aux zones d'attente un niveau de demande qui les caractériserait et qui évoluerait au cours de la journée.

Un taxi libre pourrait ainsi choisir le poste ou la zone d'attente la plus proche, mais aussi, où le niveau de demande est le plus élevé. Pour éviter que plusieurs taxis fassent le choix du même

poste ou de la même zone, le niveau de demande de chaque poste ou zone varierait en fonction du nombre de taxis déjà présents sur les lieux.

6.2.3.2 **Choix de la borne de recharge**

Un taxi libre dont le niveau de charge est au-dessous du seuil **Sc1** part en direction d'une borne de rechargement. Mais laquelle ? Plusieurs critères doivent être pris en compte. Un tel taxi se verra attribuer la borne :

- où il existe un taxi chargé disponible
- qui est la plus proche
- où le temps passé dans la file d'attente est minimal
- qui recharge le plus vite.

Note : Le calcul du temps passé dans la file d'attente dépend de plusieurs facteurs : la vitesse de rechargement de la borne de recharge, mais aussi du niveau de demande à proximité de la borne.

6.2.4 **Distribution des courses par l'opérateur**

L'opérateur est entendu ici comme étant la personne responsable de la répartition des courses entre les véhicules disponibles du réseau. Un tel opérateur peut être une machine informatique supervisée par des êtres humains. Néanmoins il doit servir à deux tâches : à traiter l'ordre de priorité des réservations de courses et la répartition des courses entre les véhicules.

6.2.4.1 **Ordre de priorité du traitement des réservations**

Dès qu'une requête a été envoyée à l'opérateur, ce dernier vérifie s'il peut mettre un véhicule à disposition de l'utilisateur.

S'il existe un véhicule libre, ou en voie de l'être, en deçà d'un certain de temps de déplacement - *disons 5 min* - celui-ci est mis à disposition de l'utilisateur. Si aucun véhicule n'est libre en deçà de ce temps, la demande de l'utilisateur est mise sur la liste d'attente du véhicule qui sera en mesure de le

servir au plus vite. Si ce véhicule met moins d'un certain temps pour parvenir jusqu'à l'usager - *ici supposons ce temps de 10 min* - la course sera réalisée, sinon elle est abandonnée.

Les demandes peuvent être rangées sur la liste d'attente par ordre d'arrivée dans le temps. Ainsi la demande faite le plus tôt sera priorisée devant la demande qui aura été faite plus tard. Cet ordre de priorité assure à l'usager d'être servi même si d'autres usagers, avantageusement plus proches du taxi, arrivent après lui. La limite de cet ordre est qu'elle ne minimise pas le nombre de kilomètres à vide.

6.2.4.2 **Ordre de priorité du traitement des taxis**

Dès qu'une requête a été envoyée à l'opérateur, celui-ci doit la satisfaire en lui envoyant un taxi électrique. Il s'agit de définir un *ordre de priorité* des véhicules à envoyer. Quels seraient les critères de choix du véhicule ?

L'opérateur doit donner la priorité au premier véhicule qui peut réaliser la course. Mais pas seulement, car une fois que le taxi est sur le lieu de rendez-vous avec l'usager il doit encore avoir à disposition suffisamment d'énergie emmagasinée dans sa batterie pour (1) réaliser le déplacement de son client et (2) réaliser un déplacement vers une borne de recharge.

Ainsi l'opérateur choisira un taxi qui sera d'abord (1) en capacité énergétique de réaliser la nouvelle course et qui mettra ensuite (2) le moins de temps possible pour se rendre sur le lieu de rendez-vous.

Note : le taxi *libre* le plus proche n'est pas nécessairement celui qui mettra le moins de temps à arriver sur le lieu de rendez-vous avec l'usager. En effet peut-être existe-t-il un taxi *occupé*, mais bientôt *libre*, qui pourrait se rendre sur les lieux avant tous les autres taxis. Ainsi le calcul du temps nécessaire à un taxi pour se rendre à l'origine d'un nouvel usager doit prendre en compte l'origine des taxis libres et occupés. De plus si nous considérons la possibilité de faire du taxi partage, alors un taxi occupé pourra s'arrêter pour prendre un deuxième usager dont l'origine et la destination sont, par exemple, sur le chemin du premier usager.

CHAPITRE 7 CONCLUSION

Ce projet de recherche est consacré à l'étude de l'offre et de la demande d'un système de transport par taxi électrique, celui de Téo Taxi qui est un nouvel opérateur à Montréal depuis décembre 2015.

7.1 Contributions

Une revue de littérature sur le taxi en général et le taxi électrique en particulier a été effectuée. Des méthodologies de prétraitement de données et un développement d'indicateurs ont été élaborées. Ceci a permis de faire des analyses utilisant les indicateurs développés. Elles portent sur l'étude de l'offre et de la demande du réseau de Téo Taxi durant les 6 derniers mois de l'année 2016. Un travail de recherche a été exposé dans le but d'intégrer la donnée spatiale à l'identification de jours-types de demande. Enfin un algorithme de génération de l'offre a été proposé, ce travail permet notamment d'avoir de manière condensée un aperçu global du fonctionnement d'un réseau de taxis électriques.

La recherche ciblant les systèmes de transport par taxi commence à s'intensifier depuis l'émergence des nouvelles technologies de l'information, cependant l'apport de connaissances sur le fonctionnement de tels systèmes reste encore en chantier. Dans ce contexte, la revue de littérature effectuée dans ce projet apporte un nouvel éclairage au lecteur qui s'intéresse aux travaux en cours dans l'industrie du taxi. Elle propose une définition d'un système de transport par taxi et fait l'état des lieux des politiques de législations encadrant son industrie. Elle propose un résumé de l'évolution historique du concept même de *taxi* et décrit les enjeux spécifiques aux taxis électriques. Enfin, une liste des indicateurs d'offre et de demande en taxi utilisés dans la littérature a été faite.

La connaissance d'un système de transport par taxi électrique passe par sa description. Comment faire un portrait global de l'utilisation de l'offre en taxi électrique ? Des outils descriptifs ont été développés, ce sont des indicateurs de description et d'aide à l'évaluation de la performance. Un certain nombre d'indicateurs sont communs avec le projet de Lacombe (2016) qui a travaillé sur la base de données de Taxi Diamond et une autre partie est spécifique à la base de données de Téo Taxi. Tous les indicateurs sont néanmoins adaptés à cette dernière. Ce projet s'inscrit dans le cadre d'un projet plus vaste ayant pour objectif d'établir un portrait global d'utilisation de l'offre

en taxi de toutes les compagnies de l'île de Montréal. Les indicateurs développés ici pour Téo Taxi sont automatisés, reproductibles et donc applicables à d'autres bases de données. Avant d'élaborer les indicateurs, une méthodologie reproductible de prétraitement de données a été effectuée. Deux défis ont été soulevés : l'automatisation d'un prétraitement de données massives et l'ajout d'un identifiant chauffeur à la base de données GPS. L'aboutissement de ce travail méthodologique de prétraitement de données et de développement d'indicateurs a conduit à un portrait global de l'utilisation de l'offre en taxis électriques de Téo Taxi sur l'île de Montréal. Cette analyse est la première de cette ampleur à avoir été faite pour le réseau de taxis électriques de Téo Taxi.

Par ailleurs, afin d'harmoniser l'offre avec la demande un travail de recherche a été effectué sur l'identification de jours-types de demande intégrant la donnée spatiale. Si un opérateur connaît les jours-types de demande pour son système de transport, il est alors capable d'adapter son offre à la demande dans le temps et dans l'espace. Par exemple, Téo Taxi serait en mesure de décider du nombre de véhicules qui serait envoyé tel jour à tel endroit.

Enfin un algorithme de génération de l'offre donne un schéma explicatif et descriptif du fonctionnement d'un réseau de taxis électriques d'une part et, d'autre part, permet de générer un nombre de véhicules et de bornes de recharge adapté à une demande donnée.

7.2 Limitations

Un certain nombre de limitations viennent encadrer ce projet. Elles sont présentées ici.

La méthode de reconstitution de l'identification du chauffeur de la table GPS se base sur les temps d'entrée et sortie des véhicules dans la zone de recharge. Ainsi durant le laps de temps où le véhicule est à l'intérieur de la zone, il n'est pas possible de lui attribuer un chauffeur. Pour gagner en précision, il serait possible de baser la reconstitution de l'identifiant à partir des temps de connexions de rechargement du véhicule. Cependant les données de charges sont à ce jour difficilement exploitables. De manière générale, la structure des données de Téo Taxi a évolué au cours de l'année 2016. De nouvelles informations sont arrivées progressivement. Ainsi les courses hélées n'ont été autorisées qu'à la fin de l'année 2016 et l'analyse des types de courses n'a pas pu être complète. Du fait de l'évolution continuelle de l'offre, car Téo Taxi part à la conquête de l'île de Montréal, une analyse du réseau de taxis électriques effectuée pour la période de

l'année 2016 ne sera pas représentative d'une année type d'opération. Il faudra pour cela attendre que l'offre se stabilise. Par ailleurs, les indicateurs développés servent à une analyse globale d'utilisation du réseau de Téo Taxi. Celle-ci peut donc être approfondie en menant des réflexions plus spécifiques sur tel ou tel objet du réseau.

L'algorithme de génération de l'offre est un travail de préparation. Des réflexions plus poussées doivent être effectuées sur les modèles de priorité du traitement des requêtes, de la répartition des courses et sur l'ordre des véhicules électriques dans la file d'attente aux bornes de recharge.

7.3 Perspectives

Ce projet a exploité les données du système de transport par taxi électrique transmises par Téo Taxi. Il s'inscrit dans un projet plus vaste qui cherche à avoir un portrait global d'utilisation du service de transport par taxi offert à Montréal. Les résultats de ce projet pourront servir à comparer l'utilisation du service Téo Taxi avec l'utilisation des services de Taxi Diamond, Taxi Hochelaga ou encore de Co-Op Taxi de l'Ouest.

Avec des données de charges structurées, l'autonomie des véhicules électriques pourrait être étudiée. En effet, l'autonomie est le défi majeur du moment pour un système de transport alimenté à l'électricité. Quelles variables diminuent ou augmentent l'autonomie des véhicules ? Où placer les bornes de recharge pour maximiser leur autonomie sur les routes et ainsi rester le plus longtemps possible disponible pour les usagers du taxi ? L'interchangement (swapping) des batteries est-il le moyen le plus efficace d'utiliser les bornes ou vaut-il mieux que les chauffeurs de taxi attendent à la borne que leurs véhicules soient rechargés ? Tels sont les questions auxquelles une analyse spécifiquement orientée sur les données de charges pourrait répondre.

L'algorithme de génération de l'offre prépare un travail de réflexion relatif à la modélisation numérique de l'offre et de la demande en transport électrique. Plusieurs scénarios pourront donc être simulés en faisant varier les paramètres influençant l'offre et la demande. À titre d'exemple, il sera possible d'observer les impacts, sur le service des taxis électriques, d'un grand nombre de bornes de recharge disponibles, ou bien d'une demande concentrée ou dispersée sur le territoire. Enfin en lien avec le taxi, de futurs projets pourraient s'intéresser à une étude portant sur les questions suivantes : est-ce que le système de transport de taxis électriques, ou de taxis en général, est une offre de transport concurrente ou complémentaire aux transports en commun, aux

véhicules particuliers, au vélopartage, etc. ? Du point de vue de la multimodalité, la contribution de ce projet pourrait aussi servir à observer à quelle part de la demande en déplacement répond Téo Taxi.

BIBLIOGRAPHIE

- Bekken, & Jon-Terje. (2007). Experiences with (De-)Regulation in the European Taxi Industry. *Transportation Research Board*, pp 31-58.
- Bischoff, J., Maciejewski, M., & Sohr, A. (2015). *Analysis of Berlin's taxi services by exploring GPS traces*. Paper presented at the Models and Technologies for Intelligent Transportation Systems (MT-ITS), 2015 International Conference on.
- Cairns, R. D., & Liston-Heyes, C. (1996). Competition and regulation in the taxi industry. *Journal of Public Economics*, 59, 1-15.
- Chang, H.-w., Tai, Y.-c., & Hsu, J. Y.-j. (2010). Context-aware taxi demand hotspots prediction. *International Journal of Business Intelligence and Data Mining*, 5(1).
- Chen, T. D., Kockelman, K. M., & Hanna, J. P. (2016). Operations of a Shared, Autonomous, Electric Vehicle Fleet : Implications of Vehicles & Charging Infrastructure Decisions. *Transportation Research Part A: Policy and Practice*, 94, 243–254.
- Damas, L. M.-M. J. G. M. F. J. M.-M. L. (2012). Online Predictive Model for Taxi Services. *LNCS*, 7619, 230–240.
- Darbéra, R. (2007). When regulators acknowledges the existence of two distinct markets for taxi services. Retrieved from <https://books.google.ca/books?hl=fr&lr=&id=rcLYAgAAQBAJ&oi=fnd&pg=PA117&dq=When+the+regulator+acknowledges+the+existence+of+two+distinct+markets+for+taxi+services.&ots=zxd4faPRM4&sig=W5-ApiYKogVmUR-CdUpIMn6nNZ4#v=onepage&q=When%20the%20regulator%20acknowledges%20the%20existence%20of%20two%20distinct%20markets%20for%20taxi%20services.&f=false>
- Darbéra, R. (2010). Transports publics et taxis : concurrence ou complémentarité ? *Ville, Rail et Transport*, pp.34-39.
- Darbéra, R. (2014a). Conflit taxis-VTC : A quoi sert la loi Thévenoud *Infrastructures & Mobilité I&M*(140).
- Darbéra, R. (2014b). VTC & PHV, une étude comparative de la manière dont les Anglais régulent l'offre de transports particuliers de personnes. *Transport*.
- Denis, Z. K. K., & Huy, F. J. S. C. V. (2017). *Predicting taxi demand at high spatial resolution: Approaching the limit of predictability*. Paper presented at the Proceedings - 2016 IEEE International Conference on Big Data.
- Dyke, K. J., Schofield, N., Barnes, M., & Member, S. (2010). *The Impact of Transport Electrification on Electrical Networks*. Paper presented at the IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS.
- Ferreira, N., Poco, J., Vo, H. T., Freire, J., & Silva, C. T. (2013). Visual exploration of big spatio-temporal urban data: A study of new york city taxi trips. *Visualization and Computer Graphics, IEEE Transactions on*, 19(12), 2149-2158. doi:10.1109/TVCG.2013.226

- Ganti, R., Mohomed, I., Raghavendra, R., & Ranganathan, A. (2011). Analysis of data from a taxi cab participatory sensor network *Mobile and ubiquitous systems: Computing, networking, and services* (pp. 197-208): Springer.
- Geradin, D. (2015). Should Uber be allowed to compete in Europe? And if so how ? *George Mason Law & Economics Research Paper*(15-29).
- Gouvernement du Québec. (2016). Loi concernant les services de transport par taxi. Retrieved from <http://www.legisquebec.gouv.qc.ca/fr/showdoc/cs/S-6.01>
- Harding, S., Kandlikar, M., & Gulati, S. (2016). Taxi apps, regulation, and the market for taxi journeys. *Transportation Research Part A*, 88, 15–25.
- Hawkins, T. R., Singh, B., Majeau-Bettez, G., & Strømman, A. H. (2013). Comparative Environmental Life Cycle Assessment of Conventional and Electric Vehicles. *Journal of Industrial Ecology*.
- Jain, A. K. (2009). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31, 651–666.
- Joonho, K., Daejin, K., Daisik, N., & Taekyung, L. (2017). Determining locations of charging stations for electric taxis using taxi operation data. *Transportation Planning and Technology*, 40, 420-433.
- JournalDeMontréal. (2016). Téo Taxi étend son service à l'aéroport Pierre-Elliott-Trudeau. Retrieved from <http://www.journaldemontreal.com/2016/06/08/teo-taxi-etend-son-service-a-laeroport-pierre-elliott-trudeau>
- Jun, L. (2013). The Multidimensional Displaying and Analysing of Taxi meter Data Based on BI. *ICTIS 2013*, pp.348-353. doi:10.1061/9780784413036.048
- King, D. A., Peters, J. R., & Daus, M. W. (2012). *Taxicabs for Improved Urban Mobility: Are We Missing an Opportunity?* Paper presented at the Transportation Research Board 91st Annual Meeting.
- Lacombe, A. (2016). *Méthodologie d'analyse et de suivi d'un système de transport par taxi.* (MAÎTRISE ÈS SCIENCES APPLIQUÉES), École Polytechnique de Montréal.
- LaPresse. (2015). L'automobile en libre-service arrive à Québec. Retrieved from <http://www.lapresse.ca/le-soleil/actualites/transports/201506/10/01-4877051-lautomobile-en-libre-service-arrive-a-quebec.php>
- Lee, J., & Park, G.-L. (2012). Demand Forecast for Electric Vehicle Sharing Systems Using Movement History Archive. *CCIS*, 341, 116–121.
- Lisa Rayle, Danielle Dai, Nelson Chan, Robert Cervero, & Shaheen, S. (2016). Just a better taxi? A survey-based comparison of taxis, transit, and ridesourcing services in San Francisco. *Transport Policy*(2016), 168–178.
- Lu, J.-L., Yeh, M.-Y., Hsu, Y.-C., Yang, S.-N., Gan, C.-H., & Chen, M.-S. (2012). *Operating Electric Taxi Fleets: A New Dispatching Strategy with Charging Plans.* Paper presented at the Electric Vehicle Conference (IEVC), 2012 IEEE International.
- Maciejewski, J. B. M. (2014). Agent-based simulation of electric taxicab fleets. *Transportation Research Procedia*, 4, 191 – 198.

- Maciejewski, J. B. M. (2016). Simulation of city-wide replacement of private cars with autonomous taxis in Berlin. *Procedia Computer Science*, 83, 237 – 244.
- Maia, R., Silva, M., Araujo, R., & Nunes, U. (2011). *Electric Vehicle Simulator for Energy Consumption Studies in Electric Mobility Systems*. Paper presented at the 2011 IEEE Forum on Integrated and Sustainable Transportation Systems.
- Meng Li, Y. J., Zuojun Shen, Fang He. (2016). Improving the electrification rate of the vehicle miles traveled in Beijing: A data-driven approach. *Transportation Research Part A*, 97, 106–120.
- Meunier, B. G. F. (2012). *Locating charging points for an electric fleet*. Paper presented at the 9th International Conference of Modeling, Optimization and Simulation, France.
- Meunier, B. G. F. (2015). Design and operation for an electric taxi fleet. *OR Spectrum*, 37, 171–194.
- Mierlo, A. N. M. M. A.-M. T. M. L. S. J. V. (2013). Environmental impacts of hybrid, plug-in hybrid, and battery electric vehicles—what can we learn from life cycle assessment? *Int J Life Cycle Assess*, 19, 1866–1890.
- Montréal, V. d. (2014). Véhicules en libre-service (VLS) : une nouvelle offre de mobilité à encadrer.
- Moreira-Matias, L., Gama, J., Ferreira, M., Mendes-Moreira, J., & Damas, L. (2013). On Predicting the Taxi-Passenger Demand: A Real-Time Approach. *LNAI*, 8154, 54–65.
- Morency, C., Verreault, H., & Demers, M. (2015). Identification of the minimum size of the shared-car fleet required to satisfy car-driving trips in Montreal. *Transportation*, 42, 435–447.
- New York City Taxi & Limousine Commission. (2014). *2014 Taxicab Fact Book*. Retrieved from http://www.nyc.gov/html/tlc/downloads/pdf/2014_taxicab_fact_book.pdf
- Park, J. J. J. Y. J. C. R. J. J. Y. (2014). Stochastic dynamic itinerary interception refueling location problem with queue delay for electric taxi charging stations. *Transportation Research Part C*, 40, 123–142.
- Pelé, N., & Morency, C. (2014a). *When, Where and How Taxis Are Used in Montreal*. Paper presented at the Transportation Research Board 93rd Annual Meeting.
- Pelé, N., & Morency, C. (2014b). When, where and how taxis are used in Montreal.
- RadioCanada. (2014). L'application de covoiturage UberX arrivée à Montréal. Retrieved from <http://ici.radio-canada.ca/nouvelle/691397/application-uberx-conducteurs-covoiturage-taxi>
- Ray, D. C. J. M. (2010). Transport and Society : Taxi! Urban Economies and the Social and Transport Impacts of the Taxicab. Retrieved from <http://site.ebrary.com/lib/polymtl/reader.action?docID=10362156>
- Rigole, P.-J. (2014). *Study of a Shared Autonomous Vehicles Based Mobility Solution in Stockholm*. (Master of Science Thesis), Royal Institute of Technology.
- Rocca, C. R. L. (2016). *Algorithmes d'optimisation pour la gestion d'une flotte de taxis électriques*. (M. Sc.), HEC Montréal.

- Roy, J. V., Leemput, N., Breucker, S. D., Geth, F., Tant, P., & Driesen, J. (2011). *An Availability Analysis and Energy Consumption Model for a Flemish Fleet of Electric Vehicles*. Paper presented at the European Electric Vehicle Congress.
- Salanova, J. M., Estrada, M., Aifadopoulou, G., & Mitsakis, E. (2011). A review of the modeling of taxi services. *Procedia Social and Behavioral Sciences*, 20, 150–161.
- Savage, T. H., & Vo, H. T. (2013). *Yellow cabs as red corpuscles*. Paper presented at the Big Data, 2013 IEEE International Conference on.
- Schaller, B. (2007). Entry controls in taxi regulation: Implications of US and Canadian experience for taxi regulation and deregulation. *Transport Policy*(14), 490–506.
- Tang, Y. M. R.-Y. K. R. H. B.-J. (2016). The analysis of the battery electric vehicle's potentiality of environmental effect : A case study of Beijing from 2016 to 2020. *Journal of Cleaner Production*, 145, 395-406.
- Taxi De France. (2017). Historique du taxi, 20e siècle. Retrieved from <https://www.taxis-de-france.com/historique/20.htm>
- Taxi Lille. (2017). L'histoire du taxi en France. Retrieved from <http://www.navette-lille.fr/historique-du-taxi-francais.html>
- Téo Taxi. (2017). Page media de Téo Taxi. Retrieved from <http://teomtl.com/medias/>
- The WashingtonPost. (2014). Taxi medallions have been the best investment in America for years. Now Uber may be changing that. Retrieved from https://www.washingtonpost.com/news/wonk/wp/2014/06/20/taxi-medallions-have-been-the-best-investment-in-america-for-years-now-uber-may-be-changing-that/?utm_term=.c773dc5385f2
- Trépanier, M. (1999). *Modélisation totalement désagrégée et orientée-objet appliquée aux transports urbains* (Ph D), Ecole Polytechnique de Montréal.
- Ville de Montréal. (2014). *Véhicules en libre-service (VLS) : une nouvelle offre de mobilité à encadrer*. Retrieved from
- Vuchic, V. R. (2007). *Urban Transit - System and Technology*.
- Yu (Marco) Nie. (2017). How can the taxi industry survive the tide of ridesourcing? Evidence from Shenzhen, China. *Transportation Research, Part C*, 242–256.
- Yunxuan, L., Jian, L., Lin, Z., & Yi, Z. (2017). Taxi booking mobile app order demand prediction based on short-term traffic forecasting. *Transportation Research Record*, 2634, 57-68.
- Zegras, D. A. C. (2011). The Taxicab as Public Transportation in Boston.
- Zhang, D.-Z., Peng, Z.-R., & Sun, D. J. (2014). *A Comprehensive Taxi Assessment Index Using Floating Car Data*. Paper presented at the Transportation Research Board 93rd Annual Meeting.
- Zhang, L., Ahmadi, M., Pan, J., & Chang, L. (2012). *Metropolitan-scale taxicab mobility modeling*. Paper presented at the Global Communications Conference (GLOBECOM), 2012 IEEE.

- Zhang, L., Chen, C., Wang, Y., & Guan, X. (2016). *Exploiting Taxi Demand Hotspots Based on Vehicular Big Data Analytics*. Paper presented at the Vehicular Technology Conference (VTC-Fall), 2016 IEEE 84th, Montréal.
- Zhang, Y. (2014). *How Do Taxis Work in Beijing? An Exploratory Study of Spatio-Temporal Taxi Travel Pattern Using GPS Data*. (Master of Urban and Regional Planning), University of California.